

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

LOW
071344258

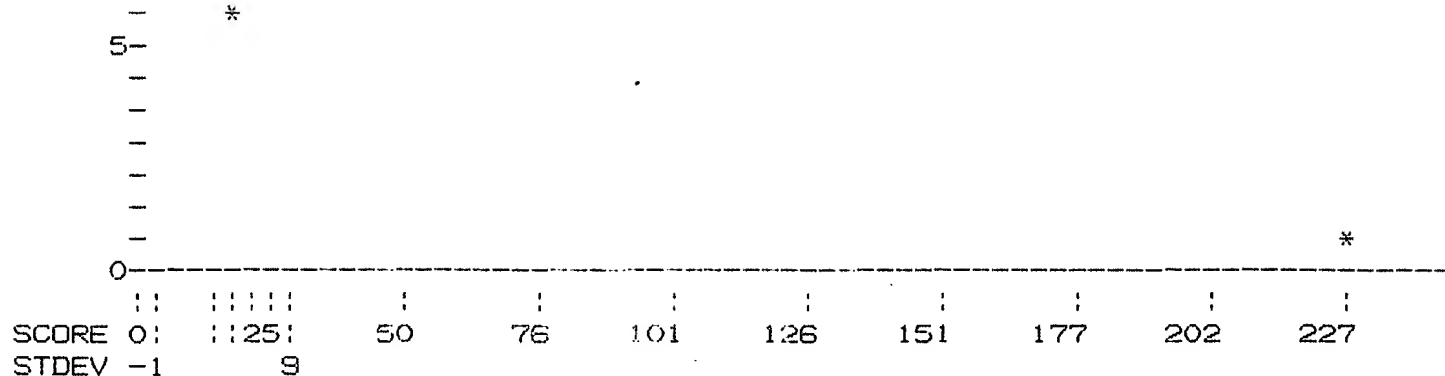
Results file low344-fig1-pir.res made by maryh on Wed 17 Apr 91 11:19:26-PDT.

Query sequence being compared: LOW344-FIG1.PEP
Number of sequences searched: 17731
Number of scores above cutoff: 3826

Results of the initial comparison of LOW344-FIG1.PEP with:
Data bank : PIR 25.0, all entries

100000-
-
N -
U50000-
M -
B -
E -
R -
-
D -*
F10000-
-
S -
E 5000-
Q - *
U -
E -
N -
C -
E *
S 1000-
- *
500-
-
-
-
-
-
100-
-
-
50- *
-
-
-
-
10-
-

-CTGGCCTT-GCTCGAGGC--GATCGACUTGGCGCTGTGCCACAGCC----CGGCCTTGCGCCAGG-GTTGG
 550 560 570 580 590 600
 ----TGTTAAAGGCGGCGCGCCCGGGCCAGCTTGGCCACCGGTGTTCCGGTAACCACTCACACGGCAGCAAGT
 610 620 630 640 650 660
 GCGCGGATCAAGGCG--CGGTGCGCCGAGGTCGGGCTGGCGCGTGGCGCCCTATTA-TCCAGCGTCGC-GTT
 670
 670 680 690 700 710 720 730
 CAGCGC--GATGGTGAGCGAGGCAAG-GCCGC-CATTTTTGAGTCCGAAGCTTGAGCCCTCACGGGTTTGTG-TA
 680 690 700 710 720 730 740
 GAGCGCCGGGCGTTTCGGCGCAGCGCGCGCAGCACGGGGTTGGG--CGAGGACGGGGTTTCGCAACAATGTGATG
 740 750 760 770 780 790
 TTGGTCAC--AGCGATGATACTG-ACGATTTGAGCTATC--TC-ACCGCCCTGCTGCGCGG-ATA-CCTCA
 750 760 770 780 790 800 810
 GCGGTAACGCTGGCCTGGCGCCTGTTGACTCGGGCGCCCGTTTCGGCCGCCCTGC-GGGCGGCACAGGCGCA
 800 810 820 830 840 850 860
 TC-GGTCTAGACCACATCCCGCACAGTGGGATTGGT-CTAGAAGATAATGCGAGTGCATCACCGCTCCTGGG
 820 830 840 850 860 870
 ACTGGACGAGGCCGC-GCAGGCCTATGGCG--CGGTGCTGCAGGACAA-GC-TG-GCCGAAGTGGTC---GG
 870 880 890 900 910 920 930
 CATCCGTT-CGTGGCAAAACACGGGCTCTCTTGATCAAGGCGCT-CATCGACCAAGGCTACATGAAC--AAA
 880 890 900 910 920 930
 CGCCTATTACGAGGCGGCCAC-GGCGC---GGC---AGGCGCTGCAT--ACTGCGG-TGAAGACACGGAGA
 940 950 960 970 980 990 1000
 TC-CTCG--TTTC-GAATGACTGGCTGTTCG-GGTTTTCEAGCTATG--TCACCAACATCATGGACGTGATG
 940 950 960 970 980 990 1000
 TCGCCCGGCGTTTCGGCCAGCATCGCGCGCGCGCGCGCGCGC-AGGCCTGGACAGCCACGGCGATGTGCTG
 1010 1020 1030 1040 1050 1060 1070
 GAT-CGCGTGAACCCC--GA-CGGGATGGCCT---TCATTC-CACTGAGAGTGATCCCATTCTACGAGAGAA
 1010 1020 1030 1040 1050 1060 1070
 CATGCGCAGGCTGCCCTGGAGCGCGCGCGCCTGGCGCAGGCGCAGGCCGAAGGCGCGCAGGC-ACGCGCGCT
 1070 1080 1090 1100 1110 1120 1130
 GGGCGTCCCACAGGAACCGCTGCCAGGCATCACTGTGACTAACC CGCGCGGTTCTGTGTACCGACTTGCC
 1080 1090 1100 1110 1120 1130 1140
 GGCCGGC---CTGGCGCAGGTCCTGGGCGTGGATCCGGC-CACGCCGATCG--TC-CTGGCGCCGGGT--CC
 1140 1150 1160 1170 1180 1190 1200
 GTGCATGACGC--CATCTGGATCCTTCCAC-GCAGCGGCCACTATTCCCGGTCAAGATACCGAACGATGAAG
 1150 1160 1170 1180 1190 1200
 G---CTGGCGCCGCAAC-GGATCGAAGATGGGAGCTGGCCCAAT--GGCTGCGGGA---CGCCCG--GCAG
 1210 1220 1230 1240 1250 1260 1270
 TCGCGCATCGATCGAT--AGGCATCTTCAATGTGATCAGG--GCTGCCACCTCCAAAGCCG--GTGGCCACC
 1210 1220 1230 1240 1250 1260 1270
 -CG-TCATCCCGCGATCAAGGCAGC-GCAGCGGGCCTGGCAGCGGCCACCGCCAGGTGATGTGGCGCGG
 1280 1290 1300 1310 1320 X
 CCTGTCGATAGTCTTGAGGCGCGGTAGGACCGTGC--TTTTCGTGAACTGCAG



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	6		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	6	7	2.65

Times:	CFU	Total Elapsed
	00:03:03.09	00:09:14.00

Number of residues: 4998063

Number of sequences searched: 17731

Number of scores above cutoff: 3826

Cut-off raised to 7.

Cut-off raised to 8.

Cut-off raised to 9.

The scores below are sorted by initial score.

Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 83 standard deviations above mean ****						
1. A28214	Phosphotriesterase - Pseudomon	325	227	251	83.28	0
**** 4 standard deviations above mean ****						
2. VGBEGX	Secreted glycoprotein gX - Pse	498	19	47	4.90	0
3. SYBSYX	Tyrosine--tRNA ligase - Bacill	419	19	43	4.90	0
4. NOHUG	Endolase gamma - Human #EC-numb	434	18	62	4.52	0
5. A27124	K+-transporting ATPase - Leish	974	18	68	4.52	0
6. S02077	Endolase gamma - Human (fragmen	433	18	62	4.52	0
7. A24742	Endolase gamma chain - Rat #EC-	434	18	63	4.52	0
8. A24405	Ice nucleation protein - Pseud	1200	17	63	4.15	0
9. QDBP4L	Hypothetical protein D-206 - B	206	17	36	4.15	0
10. A28852	Histone H3(1) - Tetrahymena py	135	17	27	4.15	0
11. B24255	Chorion class B protein L12 pr	132	17	28	4.15	0

**** 3 standard deviations above mean ****							
12.	S01921	Hypothetical protein 1 - Chlam	451	16	19	3.77	0
13.	S01022	Hypothetical protein P-2 - Chl	86	16	19	3.77	0
14.	HSXL32	Histone H3.2 - African clawed	135	16	30	3.77	0
15.	S03605	Surface glycoprotein CD14 prec	366	16	61	3.77	0
16.	A24225	Transducin beta chain - Bovine	340	16	40	3.77	0
17.	A24853	Transducin beta chain, liver -	340	16	40	3.77	0
18.	A26066	Segmentation protein eve - Fru	376	16	35	3.77	0
19.	A25457	Transducin beta chain - Bovine	340	16	40	3.77	0
20.	DAHUAL	Arachidonate 5-lipoxygenase -	674	16	65	3.77	0

The scores below are sorted by optimized score.
Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 74 standard deviations above mean ****						
1. A28214	Phosphotriesterase - Pseudomon	325	227	251	74.70	0
**** 5 standard deviations above mean ****						
2. WMBEBH	72K protein - Bovine herpesvir	664	11	72	5.42	0
**** 4 standard deviations above mean ****						
3. QQBEE3	HHLF1 protein - Cytomegaloviru	788	9	70	4.64	0
4. QQBEC3	HORF1 protein - Cytomegaloviru	846	9	69	4.26	0
5. A25902	65K protein antigen - Mycobact	588	15	69	4.26	0
6. S05506	Phosphoenolpyruvate carboxylas	966	8	69	4.26	0
7. S00893	Adenylate cyclase precursor -	1706	10	69	4.26	0
8. S02389	Cyclolysin - Bordetella pertus	1706	10	69	4.26	0
**** 3 standard deviations above mean ****						
9. A27124	H+-transporting ATPase - Leish	974	18	68	3.87	0
10. WFHUM	Mullerian inhibiting factor pr	560	13	68	3.87	0
11. GNNY5P	Genome polyprotein - Polioviru	2207	13	68	3.87	0
12. S04255	Regulatory protein qa-1S - Neu	918	8	68	3.87	0
13. GNWVWV	Genome polyprotein - West Nile	3430	10	68	3.87	0
14. QQBE8	Hypothetical BPLF1 protein - E	3149	10	67	3.48	0
15. B28894	Myeloperoxidase H7 - Human	830	11	67	3.48	0
16. OKBOG	Protein kinase, cCMP-dependent	670	11	67	3.48	0
17. DCZYPC	Pyruvate decarboxylase - Zymom	559	8	67	3.48	0
18. S02386	cyab protein - Bordetella pert	712	10	67	3.48	0
19. VGBEPB	Glycoprotein gIII precursor -	479	9	67	3.48	0
20. S00896	Ferredoxin--nitrite reductase	594	8	67	3.48	0

1. LOW344-FIG1.PEP

A28214 Phosphotriesterase - Pseudomonas diminuta MG plasm

ENTRY A28214 #Type Protein
TITLE Phosphotriesterase - Pseudomonas diminuta MG plasmid pCMS1
SOURCE Pseudomonas diminuta
ACCESSION A28214
REFERENCE (Sequence translated from the DNA sequence)
#Authors McDaniel C.S., Harper L.L., Wild J.R.
#Journal J. Bacteriol. (1988) 170:2306-2311
#Title Cloning and sequencing of a plasmid-borne gene (opd) encoding a phosphotriesterase.
SUPERFAMILY #Name phosphotriesterase
KEYWORDS esterase
COMMENT THIS SEQUENCE HAS NOT BEEN COMPARED TO THE NUCLEOTIDIC TRANSLATION.
SUMMARY #Molecular-weight 33363 #Length 325 #Checksum 539

SEQUENCE

Initial Score = 227 Optimized Score = 251 Significance = 74.70
 Residue Identity = 76% Matches = 260 Mismatches = 51
 Gaps = 28 Conservative Substitutions = 0

```

X      10      20      30      40      50      60      70
MQTRRVVLKSAAGTLLGGLAGCATWLDRAQAIGSIRARPITISEAGFTLTHEDICGSSAGFLRAWPEFFG
|||||
MQTRRVVLKSAAPRTLGGGLAGCATWLDRAQAMRSIRARPITISEAGFTLTHEDI---SAARQDSCVLGQS
X      10      20      30      40      50      60

      80      90      100      110      120      130
SRKALAEKAVRGLR-ARAAGVR---TIVD-----VSTFDIGRDVSLLAEVSRADVHIVAATGLWFDPLSM
| |
SSVAQSSSGKGCERIRAQSGWRANDICRCVDFRYRSRRQFIGR---GFAGCRR---SYLAATGLWFDPLSM
70      80      90      100      110      120      130

      140      150      160      170      180      190      200
RLRYVEELT-----QFFLREIQYGIEDTGIRAGIIKVATTGKATPFQELVLKAAARASLATGVPVTTHTAAS
|||||
RLRYVEELTLVLPVAVRFNMASKY---TGIRAGIIKVATTGKATPFQELVLKAAARASLATGVPVTTHTAAS
      140      150      160      170      180      190      200

      210      220      230      240      250      260      270
QRDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSTYLTALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSWQ
|||||
QRDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSTYLTALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSWQ
      210      220      230      240      250      260      270

      280      290      300      310      320      X 330
TRALLIKALIDQGYMKQILVSNQWLFQFSSYVTNIMQVMDRVNPDGMAFIPLRVIPFYERR
|||||
TRALLIKALIDQGYMKQILVSNQWLFQFSSYVTNIMQVMDRVNPDGMAFIH
      280      290      300      310      320      X

```

2. LOW344-FIG1.PEP

WMBEBH 72K protein - Bovine herpesvirus (type 2, strain

ENTRY WMBEBH #Type Protein
 TITLE 72K protein - Bovine herpesvirus (type 2, strain BMV)
 DATE 31-Mar-1990 #Sequence 31-Mar-1990 #Text 31-Mar-1990
 PLACEMENT 1386.0 1.0 2.0 1.0 1.0
 SOURCE bovine mamillitis virus, bovine herpesvirus 2
 ACCESSION B29242
 REFERENCE (Sequence translated from the DNA sequence)
 #Authors Hammerschmidt W., Conraths F., Mankertz J., Pauli G., Ludwig H., Buhk H. J.
 #Journal Virology (1988) 165:388-405
 #Title Conservation of a gene cluster including glycoprotein B in bovine herpesvirus type 2 (BHV-2) and herpes simplex virus type 1 (HSV-1).
 #Comment The amino acid sequence is not given in this paper.
 COMMENT The DNA sequence was obtained from GenBank, release 61.0.
 COMMENT This virus is a member of the family Herpesviridae.
 SUPERFAMILY #Name herpesvirus infected cell protein ICP18.5
 SUMMARY #Molecular-weight 72367 #Length 664 #Checksum 8190
 SEQUENCE

Initial Score = 11 Optimized Score = 72 Significance = 5.42
 Residue Identity = 26% Matches = 101 Mismatches = 203
 Gaps = 83 Conservative Substitutions = 0

```

X      10      20      30      40      50
MOTFRVVLKSAAGTLL---GGL--AGCATWLDRSAQAIGSIRARPITISEA-GFTLTHEDI
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
RLAGKICDHVTQARVRLDADEMERNLPHVVGLSEARRARALHALEVSSKMTEANSGGPAEAPGPAAAGQ---
230      X      240      250      260      270      280      290

60      70      80      90      100      110      120
CGSSAGFLRAWPEFFGSRKALAEKAVRGLR-ARAAGVRTIVDVSTFDIGRD-VSLLAEVSR----AADVHIV
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
REASA-LLDAHVFKSAPPGL--YAVSELRFWLSSGDRT--SGSTVDAFADNLSALAERERRYETGAVAVEL
300      310      320      330      340      350      360

130      140      150      160      170      180
AATG---LWFDPLSMR---LRYVEELTQFFLRRIQYGIEDTGIRAGIIKVATTGKATPFQELVL-----
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
AAFGRRGEGHFDRTFGDRVASLDMMVAL---FVGGQSAAPDDQIEALVRACYNHHL SAP----VLRQLAGSE
370      380      390      400      410      420

190      200      210      220      230      240
--KAAARASLATGVFVTTHTA-ASWRDG----ERGRPPFLSPKLEPSRVCIGHSDDTDDL SYLTALLRGYLI
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
HGDAEALRSALEG-----LHAEDPIFGDGNAEKEARRAPSL-----GGGPEDDWAALAARAAADV GARR
430      440      450      460      470      480

250      260      270      280      290      300      310
GLDHIPHSAIGLEDNASASPLLGINSWQTRALLIKALIDQGY-----MKQILVSN--DWLFGFSSYVTNIMD
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
RL-----YADRLTKRSLAS--LGRQVREGRGELEKMLRVSTYGEVLPTVFAAVCNGFAARTRFCELTARAGT
490      500      510      520      530      540

320      330      340      X
VMD-RVNPDGMAFIPLRVIPFYERRAS-HPKRCBASL
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
VIDNRGNPD--TFDTHR-----FMRASLMRHRVDFALLPGITHQFFE
550      560      570      580

```

3. LOW344-FIG1.PEP

Q0BEE3 HHLF1 protein - Cytomegalovirus (strain AD169)

```

ENTRY      Q0BEE3      #Type Protein
TITLE      HHLF1 protein - Cytomegalovirus (strain AD169)
DATE       30-Sep-1989 #Sequence 30-Sep-1989 #Text 31-Dec-1989
PLACEMENT  1358.0      1.0      1.0      2.0      1.0
SOURCE      human cytomegalovirus, human herpesvirus 5
ACCESSION  C27349
REFERENCE   (Sequence translated from the DNA sequence)
#Authors    Weston K., Barrell B.G.
#Journal     J. Mol. Biol. (1986) 192:177-208
#Title       Sequence of the short unique region, short repeats,
              and part of the long repeats of human
              cytomegalovirus.
COMMENT     The DNA sequence was obtained from EMBL, release 13.
COMMENT     This virus is a member of the family Herpesviridae.
GENETIC
#Name        HHLF1
SUPERFAMILY #Name cytomegalovirus HHLF1 protein
SUMMARY     #Molecular-weight 83881 #Length 788 #Checksum 7858
SEQUENCE

```

```

Initial Score      =      9      Optimized Score      =      70      Significance      =      4.64
Residue Identity   =      23%      Matches              =      90      Mismatches       =      241
Gaps               =      55      Conservative Substitutions      =      0

```

```

X      10      20      30      40      50
MOTFRVVLKSAAGTLLLEGLAGCATWLDRSAQAIG-----SIRARPITISEAGFTLTHEDI

```

```

ASAPHPASLLTAVRRHILNERLCCGWLALGAVLPARWLGCAAGPATGTAAGTTSPPAASGTETEAAGGDAPCA
330 X      340      350      360      370      380      390

60      70      80      90      100      110      120
ICG--SSAGFLRAWPEFFGSKA-----LA-LAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRADVH
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
IAGAVGSAVPVPIPEPYGAAGGGAICVPNADAHAVVSEADAAAAAAPTVMVGSTAMAGPAAS--GTVPRAMLVV
400      410      420      430      440      450      460

130      140      150      160      170      180
IVAATGLWF-DPPLSMRLRYVEELTQFFLREIDYGIEDTG-IRAGIIKVATTGKATPFQE-----LVLKA
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
LLDELGAVFGYCPLDGHVYPLAAELSHFLRAGVLGALALGRESAPAAEAARRLLPELDREQWERPRWDALHL
470      480      490      500      510      520      530

190      200      210      220      230      240      250
AARASLATGVPVTTHTAASQRDGERGRPFPLSPKLEPSRVCIGHSDDTDDLSYLTALLR--GYLIGLD--HI
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
HPRAALWAREP-HGQLAFLLRPG-RGEAEVLTATKHPAICANVEDYLQD-----ARRRADAGALGLDLATV
540      550      560      570      580      590      600

260      270      280      290      300      310
PHSAIG-----LEDNASASPIILGIRSWQTRALLIKALIDQGYMKQILVSNDWLFGFSSYVTNIMDVMD
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
VMEAGGQMIHKKTCKPKGKEDESLMKGKHSRYTR-PTEPPLTQASLGRALRRDDEDWKPS-----RLPGED
610      620      630      640      650      660

320      330      340      X
RVNPDGMAFIPL---RVIPFYERR-ASHRKRCQASL
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
SWYDLDETFWVLGSNRKNDVYQRRWKKTVLRCGLEIDRPMPTVPKG
670      680      690      700 X      710

```

4. LOW344-FIG1.PEP

Q0BEC3 H0RF1 protein - Cytomegalovirus (strain AD169)

```

ENTRY      Q0BEC3      #Type Protein
TITLE      H0RF1 protein - Cytomegalovirus (strain AD169)
DATE       30-Sep-1989 #Sequence 30-Sep-1989 #Text 30-Sep-1989
PLACEMENT  1358.0      1.0      1.0      1.0      1.0
SOURCE     human cytomegalovirus, human herpesvirus 5
ACCESSION  C26078
REFERENCE  (Sequence translated from the DNA sequence)
  #Authors  Weston K., Barrell B.G.
  #Journal  J. Mol. Biol. (1986) 192:177-208
  #Title    Sequence of the short unique region, short repeats,
            and part of the long repeats of human
            cytomegalovirus.
COMMENT    The DNA sequence was obtained from EMBL, release 13.
COMMENT    This virus is a member of the family Herpesviridae.
GENETIC
  #Name     H0RF1
SUPERFAMILY #Name cytomegalovirus H0RF1 protein
SUMMARY    #Molecular-weight 91047 #Length 846 #Checksum 2604
SEQUENCE

```

```

Initial Score      =      9      Optimized Score      =      69      Significance      =      4.26
Residue Identity   =     22%      Matches              =      86      Mismatches      =     248
Gaps               =      46      Conservative Substitutions      =      0

```

```

X      10      20      30      40      50
M0TRRVVLKSAAAGTLLGGLAGCATWLDRSA-QAIG-----SIRARPITISEAGFTLTHER
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
ASAPHPASLLTAVRRHILN0RLCCGWLALGAVLPARWLGCAAGPATGTAAGTTSPPAASGTETEAAGGDAPCA

```



```

X      10      20      30      40      50
MOTRRVVLKSAA---ADTLLGLAGCATWLDRSABAIGSIR---ARPITI---SEAGFTLTHER
|      |      |      |      |      |      |      |      |      |      |
VPGRDGETQPASCGRPSPRALHPABVEIGSGRHPVT-----LASFLIRRNFAMAKTIA YDEEARRGL--ER
      10      20      30      40      50      60

      60      70      80      90      100      110      120
ICGSSAGFLRAWFEFFGSRKALAEKAVFGLPARAAGVRTIVDVSTFD----IGRDVSL LAEVSRAADVHIVA
|      |      |      |      |      |      |      |      |      |      |
GLNSLADAVKVTLGPKG-FNVLEKFWLAPTITNDGVSIAKEIELEDPYEKIG--AELVKEV--AKKTDDVA
      70      80      90      100      110      120      130

      130      140      150      160      170      180      190
ATGLWFDPLSLMPLRYVEELTEFFLPEFQVGIIDTQIPAGI-----IKVATT-GKATPFDELVLKAAARASLA
|      |      |      |      |      |      |      |      |      |      |

```



```

50      60      70      80      90     100     110
GFTLT HEDICG--SSAGFLRAWPEFFG-----SRKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSLAE
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
GMSLTDDAPAGQKAAGAEIALQLTGTVELASSIALA-----LAAARGVTSGLQVAGASAG-----
120      130      140      150      160      170
VSRADVHIVAATGLWFDPLSMRI.RYVEELTQFFLR EIQYGIE-----DTGIRAGIIKVATT
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
---AAAGALAAALSPMEIYGLVQSHYADQLDKLAGESSAYGYEGDALLAQLYRDKTAAEGAVAGVSAVLST
610      620      630      640      650      660      670

180      190      200      210      220      230
GKATPFQELVLKAAARASLATGVPVTTHTAASQRDG-----ERGRPPFLSPKLEPSRVCIGHSDDTDDL SYL
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
VGA-----AVSIAAA-AS-VVGAPVAVVT--SLLTGALNGILRGVQQPIIEKL-----ANDYARKIDELGGP
680      690      700      710      720      730

240      250      260      270      280      290      300
TA-----LLRGYLIGLDHIPHSNIGLEDNASASPLLGI RSWQTRALLIKALIDQGYMKQILVSNOWLFGF
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QAYFEKNLQARHEQLANS DGLRKMLADLQAGWNASSVIG---VQTEISK SAL-----ELAAITGNADNL--K
740      750      760      770      780      790

310      320      330      340      X
SSYVTNIMDVM DRVNP DGM AFI P--LRVIFPYERRASHRKRCQASL
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
SVDV-----FVDRFVQGERVAGQPVVL DVAAGGIDIAS-RKGERPALTFITPLAAPG
800      810      820      830      840

```

8. LDW344-FIG1.PEP

S02389 Cyclolysin - Bordetella pertussis

```

ENTRY          S02389      #Type Protein
TITLE          Cyclolysin - Bordetella pertussis
INCLUDES       adenylate cyclase #EC-number 4.6.1.1\ hemolysin
SOURCE         Bordetella pertussis
ACCESSION      S02389
REFERENCE
#Authors       Glaser P. , Sakamoto H. , Bellalou J. , Ullmann A. ,
                Danchin A.
#Journal       EMBO J. (1988) 7:3997-4004
#Title         Secretion of cyclolysin, the calmodulin-sensitive
                adenylate cyclase-haemolysin bifunctional protein
                of Bordetella pertussis.
#Reference-number S02389
#Accession     S02389
#Molecule-type DNA
#Residues      1-1706 (GLA)
GENETIC
#Name          cyaA
COMMENT        THIS SEQUENCE HAS NOT BEEN COMPARED TO THE
                NUCLEOTIDE TRANSLATION.
FEATURE
1-312          #Domain adenylate cyclase (ADE)\
313-1706       #Domain hemolysin (HEM)
SUMMARY        #Molecular-weight 177476 #Length 1706 #Checksum 6271
SEQUENCE

```

```

Initial Score   =      10   Optimized Score   =      69   Significance =    4.26
Residue Identity =    23%   Matches            =     93   Mismatches   =    206
Gaps           =     97   Conservative Substitutions =      0

```

```

X      10      20      30      40
MQTPRVVLKSAAGTLLGLAGCATWLDRSQAIGSIR-----ARPITI---SEA

```

```

LMTQFGRAGSTNTFGEAASLSAAVPEL-GEASDAVAETVSGFFRGSSRWAGGFGVAGGAMALGGGIAAAVGA
      490          500          510          520          530          540          550

      50          60          70          80          90          100          110
GFTLT HEDICG--SSAGFLRAWPEPEG-----SRKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSLAE
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
GMSLTDDAPAGQKAAAGAEIALQLTGGIVELASSIALA-----LAAARGVTSGLQVAGASAG-----
      560          570          580          590          600

      120          130          140          150          160          170
VSRAADVHIVAATGLWFDPPLSMRILRYVEELTQFFLREIQYGIE-----DTGIRAGIIKVATT
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
---AAAGALAAALSPMEIYGLVQQSHYADQLDKLAGQESSAYGYEGDALLAQLYRDKTAAEGAVAGVSAVLST
      610          620          630          640          650          660          670

      180          190          200          210          220          230
GKATPFQELVLKAAARASLATGVPVTTHTAASQRDG-----ERGRPPFLSPKLEPSRVCIGHSDDTDDLSTL
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
VGA-----AVSIAAA-AS-VVGAPVAVVT--SLLTGALNGILRGVQQPIIEKL-----ANDYARKIDELGGP
      680          690          700          710          720          730

      240          250          260          270          280          290          300
TA-----LLRGYLIGLDHIPHSAIGLEDNASASPLLGI RSWQTRALLIKALIDQGYMKQILVSNDFLFGF
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QAYFEKNLQARHEQLANS DGLRKMLADLQAGWNASSVIG---VQTEISK SAL-----ELAAITGNADNL--K
      740          750          760          770          780          790

      310          320          330          340          X
SSYVTNIMDVM DRVNP DGM AFI P--LRVIFFYERRASHRKRCQASL
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
SVDV-----FVDRFVQGERVAGQPVVLDVAAGGIDIAS-RKGERPALTFITPLAAPG
      800          810          820          830          840

```

9. LOW344-FIG1.PEP

A27124 H+-transporting ATPase - Leishmania donovani

```

ENTRY      A27124      #Type Protein
TITLE      H+-transporting ATPase - Leishmania donovani
            #EC-number 3.6.1.35
ALTERNATE-NAME  proton-transporting ATPase
SOURCE      Leishmania donovani
ACCESSION   A27124
REFERENCE   (Sequence translated from the DNA sequence)
            #Authors   Meade J. C. , Shaw J. , Lemaster S. , Gallagher G. ,
                        Stringer J. R.
            #Journal    Mol. Cell. Biol. (1987) 7:3937-3946
            #Title      Structure and expression of a tandem gene pair in
                        Leishmania donovani that encodes a protein
                        structurally homologous to eucaryotic
                        cation-transporting ATPases.
            #Comment    The authors translated the codon AGA for residue 352
                        as Lys.
SUMMARY     #Molecular-weight 107476 #Length 974 #Checksum 834
SEQUENCE

```

```

Initial Score      =      18      Optimized Score      =      68      Significance      =      3.87
Residue Identity   =     24%      Matches               =      93      Mismatches       =     210
Gaps               =      82      Conservative Substitutions      =      0

```

```

      X          10          20          30          40          50
MSTRR-----VVLKSAAGTLLGGLAGCATWLDRSQAIGSIRARPITISEAGFTLTHE---
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
FLDPPRPD'TKDTIRRSKEYGVQVKNHITQDPLLI AKEMC-RMLDLDPNILTADKLPQIKDANDLPEDLGEKYG
      500          510          520          530          540          550          560

```

```

      60      70      80      90     100     110     120
DICGSSAGFLRAWPEFFGSRKALAEKAVPGLRAR---AAGVRTIVDVSTFDIGRDVSLLAEVSRADVHIVA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DMMLSVGGFAQVFPE-----HKFMIVETLRWRGYTCAMTGGGVNDAPALKRADV--GIAVHGATDAARAA
570      580      590      600      610      620      630

      130     140     150     160     170
ATGLWFDPLSMRLRYVE-----ELTBFFLRREIBYGIET---GIRAGIIKVATTGKA-----TPFQ---E
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ADMVLTEPGLS---VVVEAMLVSREVFAQMLSFLTYRISATLGLVCFFFIACFSLTPKAYGSVDPHFQFFHL
      640      650      660      670      680      690      700

180     190     200     210     220     230     240
LVLKAAARASLATGVPVT---THTAASWRDGERGRFPFLSPKLEPSRVCIGHSDDTDDLSTALLRGYLIG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
PVLMFMLITLLNDGCLMTIGYDHVIPS---ERPQKWNL-PVVFVS-----ASILAAVACGSSLM
      710      720      730      740      750

250     260     270     280     290     300     310
LDHIPHSAIGLEDNASASPLLGIKSWQTR---ALLIKALIDQGYMKQILVSNWDLFGFSSYVTNIMDVMDR
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
L-----LWIGLE---GYSSQYYENSWFHRLGLAQLPQGLVTMMYLK-ISIS-DFLTIFSSRTGGHFFFYMP
      760      770      780      790      800      810

      320     330     340     X
VNP--DGMAFIPLRVIPFYERRAS--HRKRCQASL
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
PSPILFCGAIISLLV---STMAASFVHKSRPDNVLTEGLAWGQTN
820      830      840      850      860

```

10. LOW344-FIG1.PEP

WFHUM Mullerian inhibiting factor precursor - Human

```

ENTRY      WFHUM      #Type Protein
TITLE      Mullerian inhibiting factor precursor - Human
ALTERNATE-NAME  Mullerian inhibiting substance (MIS)
DATE       13-Aug-1986 #Sequence 13-Aug-1986 #Text 30-Jun-1987
PLACEMENT  596.0  5.0  1.0  1.0  1.0
SOURCE     Homo sapiens #Common-name man
ACCESSION  A01397
REFERENCE  (Sequence translated from the DNA sequence)
#Authors   Cate R.L., Mattaliano R.J., Hession C., Tizard R.,
            Farber N.M., Cheung A., Ninfa E.G., Frey A.Z.,
            Gash D.J., Chow E.P., Fisher R.A., Bertonis J.M.,
            Torres G., Wallner B.P., Ramachandran K.L., Ragin
            R.C., Mangano T.F., MacLaughlin D.T., Donahoe
            P.K.
#Journal   Cell (1986) 45:685-698
COMMENT    This protein is homologous to transforming growth
            factor beta, inhibin alpha chain, and inhibin A
            and B chains. The area of best homology
            corresponds to the mature proteins.
COMMENT    Although it does not compete with EGF for receptor
            binding sites, MIS can inhibit the
            autophosphorylation of the EGF receptor in vitro.
GENETIC
#Introns   138/1, 185/3, 222/1, 275/2
SUPERFAMILY #Name inhibin
KEYWORDS    testicular glycoprotein\ gonadal differentiation\
            antitumor agent\ Mullerian duct\ TGF-beta homolog\
            inhibin homolog
FEATURE
1-25        #Domain signal and propeptide sequence
            (SIG)

```

#Binding-site carbohydrate (Asn)
(potential)

SUMMARY

SEQUENCE

```
#Molecular-weight 59192 #Length 560 #Checksum 3812
```

Initial Score	=	13	Optimized Score	=	68	Significance	=	3.87
Residue Identity	=	24%	Matches	=	99	Mismatches	=	211
Gaps	=	92	Conservative Substitutions	=	0			

```

X      10      20      30      40      50
MQTRRVVL-KSAAAGTLLG-GLA-----GCATWL-DRSQA-IGSIRARPIT-ISEAGFTL
|| || || || | || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
LPGAQSLCPSRDTRYLVLAVD RPAGAWRGSG LALTLP RGEDSRLSTARLQALLFGDDHRCFTRMTPALLLL
190      200      210      220      230      240      250

60      70      80      90      100     110
THEDICGSSA-GFLRAWPEFFGSRKALAE----KAVRGLRARAAAGVRTIVDVSTFDIGRDVSL LAEVSRAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
PRSEPAPLPAHGQLDTPFP PPRP SAELEESPPSADPFLETLTRLR-----ALRVPPARASAPRLAL
260      270      280      290      300      310

120     130     140     150     160     170     180
DVHIVA--ATGL--WFDPLSMRLRYVEELTQFFLREIQYGI EDTGIRAGIIKVATTGKATPFQELV---LK
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DPDALAGFPQGLVNLS DPAAIERLIDGEEPLL LLLLRR---PTAATTGDPA PLHDPTSAPWATALARRVAAELQ
320     330     340     350     360     370     380

190     200     210     220     230     240
AAA--RASLATGVPTTHTAASQRDGERGRPPFLSPKLEPSRVCIGHSDDTDDL SYLTAL--LRGYLIGLD-
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AAAAELRSLPGLPPATA PLLA-----RI LLALCP-GGP-----GG LGDPLRAL LLLKALQGLRVEWRGRDP
390     400     410     420     430     440

250     260     270     280     290     300     310
HIPHSA-IGLEDNASASP-LLGISWETRA---LLI---KALIDQGYMKQILVSN DWLFGFSSYVTNIMDV
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
RGPGRAQRSAGATAADGPCALREL SVDLRAERSVLIPETYQANNCQG-----VCGWPQSDRNPRYG
450     460     470     480     490     500

320      330      340      X
MDRVNPDGM-----AF IPLRVIPFY-----ER-RASHRKRCQASL
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
NHVVLLLMQARGAALARPPCCVPTAYAGKI LISLSEERISAHHPNMVATECGCR
510      520      530      540      550      X 560

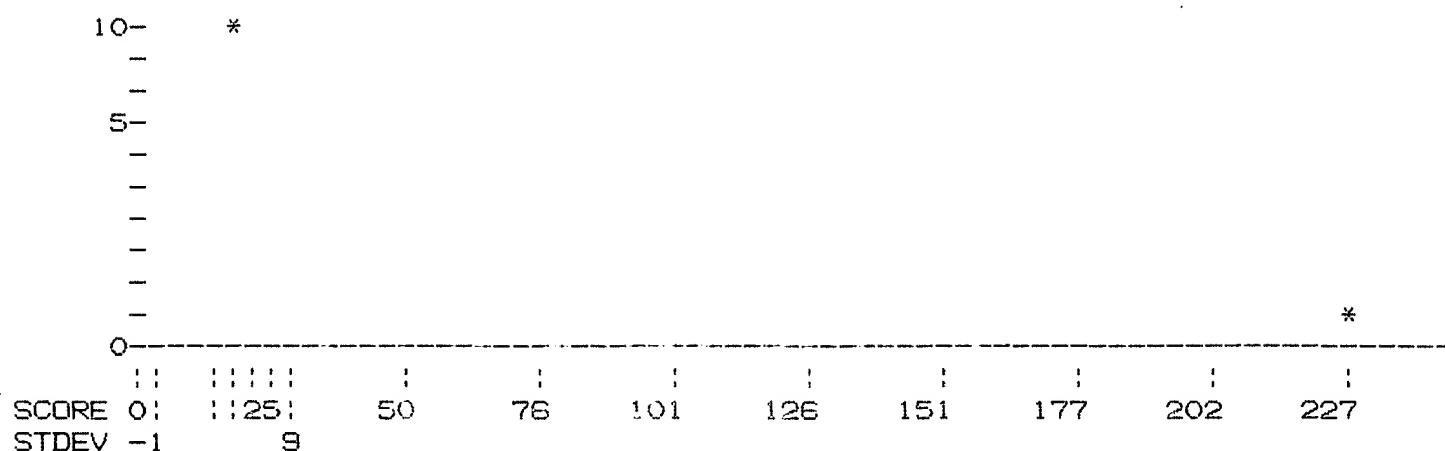
```

Results file low344-fig1-spt.res made by maryh on Wed 17 Apr 91 11:34:12-PDT.

Query sequence being compared: LOW344-FIG1.PEP
Number of sequences searched: 15409
Number of scores above cutoff: 3758

Results of the initial comparison of LOW344-FIG1.PEP with:
Data bank : Swiss-Prot 14, all entries

100000-
-
N -
U50000-
M -
B -
E -
R -
-
D -*
F10000-
-
S -
E 5000-
Q - *
U -
E -
N -
C -
E -
S 1000-
-
* *
500-
-
-
-
-
-
-
100-
-
-
50- *
-
-



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	1	Joining penalty	20
Gap penalty	1.00	Window size	32
Gap size penalty	0.05		
Cutoff score	6		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	6	7	2.69

Times:	CPU	Total Elapsed
	00:03:15.00	00:13:53.00

Number of residues:	4914263
Number of sequences searched:	15409
Number of scores above cutoff:	3758

Cut-off raised to 7.
Cut-off raised to 8.
Cut-off raised to 9.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 82 standard deviations above mean ****						
1. OPD\$PSEDI	PHOSPHOTRIESTERASE (EC 3.5.1.1)	325	227	251	82.28	0
**** 4 standard deviations above mean ****						
2. MTC3\$CHVN1	MODIFICATION METHYLASE CVIB II	377	19	25	4.84	0
3. SYY\$BACCA	TYROSYL-TRNA SYNTHETASE (EC 6.1.1.1)	419	19	43	4.84	0
4. VGLX\$PRV	SECRETED GLYCOPROTEIN GX (GENE)	498	19	47	4.84	0
5. ENOG\$RAT	GAMMA ENOLASE (EC 4.2.1.11) (2	433	18	63	4.47	0

6.	ENOG\$HUMAN	GAMMA ENOLASE (EC 4.2.1.11) (2	433	18	62	4.47	0
7.	ATXA\$LEIDO	PROBABLE E1-E2 TYPE CATION ATP	974	18	68	4.47	0
8.	ICEN\$ERWHE	ICE NUCLEATION PROTEIN (GENE N	1258	18	63	4.47	0
9.	ATXB\$LEIDO	PROBABLE E1-E2 TYPE CATION ATP	974	18	68	4.47	0
10.	H31\$TETPY	HISTONE H3.1.	135	17	27	4.10	0
11.	KCCA\$RAT	CALCIUM/CALMODULIN-DEPENDENT P	478	17	57	4.10	0
12.	KCCA\$MOUSE	CALCIUM/CALMODULIN-DEPENDENT P	478	17	57	4.10	0
13.	ICEN\$PSESY	ICE NUCLEATION PROTEIN (GENE N	1200	17	63	4.10	0
14.	CHA2\$BOMMO	CHORION CLASS A PROTEIN L12 PR	132	17	28	4.10	0
15.	Y206\$LAMBD	HYPOTHETICAL PROTEIN DRF206.	206	17	36	4.10	0
**** 3 standard deviations above mean ****							
16.	ARSA\$HUMAN	ARYLSULFATASE A PRECURSOR (EC	507	16	64	3.72	0
17.	HMEV\$DROME	SEGMENTATION PROTEIN EVEN-SKIP	376	16	35	3.72	0
18.	GBB1\$HUMAN	GUANINE NUCLEOTIDE-BINDING PRO	340	16	40	3.72	0
19.	H32\$XENLA	HISTONE H3.2.	135	16	30	3.72	0
20.	CD14\$MOUSE	CD14 DIFFERENTIATION ANTIGEN P	366	16	61	3.72	0

The scores below are sorted by optimized score.
Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 76 standard deviations above mean ****						
1. OPD\$PSEDI	PHOSPHOTRIESTERASE (EC 3.5.-.-	325	227	251	76.20	0
**** 4 standard deviations above mean ****						
2. YHL1\$HCMVA	HYPOTHETICAL PROTEIN HHLF1.	788	9	70	4.74	0
3. YQR1\$HCMVA	HYPOTHETICAL PROTEIN HQRF1.	846	9	69	4.34	0
4. CYAA\$BORPE	CALMODULIN-SENSITIVE ADENYLATE	1706	10	69	4.34	0
5. CAP1\$MESCR	PHOSPHOENOLPYRUVATE CARBOXYLAS	966	8	69	4.34	0
**** 3 standard deviations above mean ****						
6. MIS\$HUMAN	MULLERIAN INHIBITING FACTOR PR	560	13	68	3.95	0
7. ATXA\$LEIDO	PROBABLE E1-E2 TYPE CATION ATP	974	18	68	3.95	0
8. POLG\$WNV	GENOME POLYPROTEIN (CAPSID PRO	3430	10	68	3.95	0
9. ATXB\$LEIDO	PROBABLE E1-E2 TYPE CATION ATP	974	18	68	3.95	0
10. ATP0\$OENBI	ATP SYNTHASE ALPHA CHAIN, MITO	511	9	68	3.95	0
11. POLG\$POL2L	GENOME POLYPROTEIN (COAT PROTE	2207	13	68	3.95	0
12. PGKH\$WHEAT	PHOSPHOGLYCERATE KINASE, CHLOR	480	8	68	3.95	0
13. EXON\$HSV11	ALKALINE EXONUCLEASE (EC 3.1.1	626	9	67	3.55	0
14. KGP\$BOVIN	CGMP-DEPENDENT PROTEIN KINASE	670	11	67	3.55	0
15. KGPB\$HUMAN	CGMP-DEPENDENT PROTEIN KINASE,	686	11	67	3.55	0
16. VGL3\$PRV	GLYCOPROTEIN GIII PRECURSOR.	479	9	67	3.55	0
17. PHYB\$ARATH	PHYTOCHROME B (GENE NAME: PHYB	1172	9	67	3.55	0
18. ATI1\$HSV11	ALPHA TRANS-INDUCING FACTOR 73	693	9	67	3.55	0
19. PGCA\$RAT	CARTILAGE-SPECIFIC PROTEOGLYCA	2124	9	67	3.55	0
20. PYR1\$YEAST	CARBAMOYL-PHOSPHATE SYNTHETASE	1456	11	67	3.55	0

1. LOW344-FIG1. PEP

OPD\$PSEDI PHOSPHOTRIESTERASE (EC 3.5.-.-) (GENE NAME: OPD).

```

ID  OPD$PSEDI      STANDARD;      PRT;      325 AA.
AC  P13739;
DT  01-JAN-1990   (REL. 13, CREATED)
DT  01-JAN-1990   (REL. 13, LAST SEQUENCE UPDATE)
DT  01-APR-1990   (REL. 14, LAST ANNOTATION UPDATE)
DE  PHOSPHOTRIESTERASE (EC 3.5.-.-) (GENE NAME: OPD).
OS  PSEUDOMONAS DIMINUTA.
OG  PLASMID PCMS1.
OC  PROKARYOTA; BACTERIA; GRAM-NEGATIVE AEROBIC RODS AND COCCI;
OC  PSEUDOMONADACEAE.

```

```

ID      YHL1$HCMVA          STANDARD;          PRT;          788 AA.
AC      P09695;
DT      01-MAR-1989    (REL. 10, CREATED)
DT      01-MAR-1989    (REL. 10, LAST SEQUENCE UPDATE)
DT      01-JAN-1990    (REL. 13, LAST ANNOTATION UPDATE)
DE      HYPOTHETICAL PROTEIN HHLF1.
OS      HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC      VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE.
RN      [1] (SEQUENCE FROM N. A.)
RA      WESTON K., BARRELL B.G.;
RL      J. MOL. BIOL. 192:177-208(1986).
CC      -!- SIMILARITY: TO HHLF1, HHLF5, HHLF6, HHLF7, AND HQRF1.
DR      EMBL; X04630; HEHCMVU.
DR      PIR; C27348; QQBEE3.
KW      HYPOTHETICAL PROTEIN.
FT      CARBOHYD      75      76      POTENTIAL.
FT      CARBOHYD      110     118     POTENTIAL.
FT      CARBOHYD      223     223     POTENTIAL.

```

SQ SEQUENCE 788 AA; 20981 MW; 2964529 CN;

Initial Score = 9 Optimized Score = 70 Significance = 4.74
 Residue Identity = 23% Matches = 90 Mismatches = 241
 Gaps = 55 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      MSTRRVVLKSAAGTLLGGLAGCATWLD RSA-QAIG-----SIRARPITISEAGFTLTHED
      ||      ||      ||      ||      ||      ||      ||      ||      ||
ASAPHPASLLTAVRRHLNQRLLCCGWLALGAVLPARWLGCAAGPATGTAAGTTSPPAASGTETEAGGDAPCA
      330 X      340      350      360      370      380      390

      60      70      80      90      100      110      120
      ICG--SSAGFLRAWPEFFGSRKA-----LA-EKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRADVH
      ||      ||      ||      ||      ||      ||      ||      ||      ||
      IAGAVGSAPVPPQPYGAAGGGGAIQVFNADAHAVVGADAAAAAAPTVMVGSTAMAGPAAS--GTVPRAMLVV
      400      410      420      430      440      450      460

      130      140      150      160      170      180
      IVAATGLWF-DPPLSMRLRYVEELTQFFLREIQYGIEDTG-IRAGIIKVATTGKATPFQE-----LVLKA
      ||      ||      ||      ||      ||      ||      ||      ||      ||
      LLDDELGAVFGYCPLDGHVYPLAAEISHFLRAGVLGALALGRESAPAAEAARRLLPELDREQWERPRWDALHL
      470      480      490      500      510      520      530

      190      200      210      220      230      240      250
      AARASLATGVPVTTHTAASQRDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSTALLR--GYLIGLD--HI
      ||      ||      ||      ||      ||      ||      ||      ||      ||
      HPRAALWAREP-HGQLAFLLRPG-RGEAEVLTATKHFAICANVEDYLQD-----ARRRADAGALGLDLATV
      540      550      560      570      580      590      600

      260      270      280      290      300      310
      PHSAG-----LEDNASASPLLGI RSWQTRALLIKALIDQGYMKQILVSNWDLFGFSSYVTNIMDVMD
      ||      ||      ||      ||      ||      ||      ||      ||      ||
      VMEAGGQMIHKKTCKPKGKEDESLMKGKHSRYTR-PTEPPLTPQASLGRALRRDDEDWKPS-----RLPGED
      610      620      630      640      650      660

      320      330      340      X
      RVNPDGMAFIPL---RVIPFYERR-ASHRKRCQASL
      ||      ||      ||      ||
      SWYDLDETFWVLGSNRKNDVYQRRWKKTVLQGLEIDRPMPTVPKG
      670      680      690      700 X      710
  
```

3. LOW344-FIG1.PEP

YQR1\$HCMVA HYPOTHETICAL PROTEIN HQRF1.

ID YQR1\$HCMVA STANDARD; PRT; 846 AA.
 AC P09715;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-JAN-1990 (REL. 13, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL PROTEIN HQRF1.
 OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
 OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE.
 RN [1] (SEQUENCE FROM N. A.)
 RA WESTON K., BARRELL B. G.;
 RL J. MOL. BIOL. 192:177-208(1986).
 CC -!- SIMILARITY: TO HHLF1, HHLF1, HHLF5, HHLF6, AND HHLF7.
 DR EMBL; X04650; HEHCMVU.
 DR PIR; C26070; Q@BEC3.
 KW HYPOTHETICAL PROTEIN.
 FT CARBOHYD 76 76 POTENTIAL.
 FT CARBOHYD 113 118 POTENTIAL.
 FT CARBOHYD 223 223 POTENTIAL.
 SQ SEQUENCE 846 AA; 91047 MW; 3448605 CN;

Initial Score = 9 Optimized Score = 69 Significance = 4.34
 Residue Identity = 22% Matches = 86 Mismatches = 248
 Gaps = 46 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      MSTRRVVLKSAAGTLLGGLAGCATVLDRSA-QAIG-----SIRARPITISEAGFTLTHER
      ::      ::      ::      ::      ::      ::      ::      ::      ::
ASAPHPASLLTAVRRHLNQRLLCCGVLALGAVLPARWLGCAAGPATGTAAGTTSPPAASGTETEAAGGDAPCA
      330 X      340      350      360      370      380      390

      60      70      80      90      100      110      120
ICG--SSAGFLRAWPEFFGSRKA----LA-EKAVRGLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRAADVH
::      ::      ::      ::      ::      ::      ::      ::      ::
IAGAVGSAVPVPPQPYGAAGGGAICVFNADAHAVVGADAAAAAAPTVMVGSTAMAGPAAS--GTVPRAMLVV
      400      410      420      430      440      450      460

      130      140      150      160      170      180
IVAATGLWF-DPPLSMRLRYVEELTQFFLREIQYGIEDTG-IRAGIIKVATTGKATPFQE-----LVLKA
      ::      ::      ::      ::      ::      ::      ::      ::
LLDELGAVFGYCPLDGHVYPLAAELSHFLRAGVLGALALGRESAPAAEAARRLLPELDREQWERPRWDALHL
      470      480      490      500      510      520      530

      190      200      210      220      230      240      250
AARASLATGVPVTTHTAASQRDGE-RGRPFPLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIGLDHIPHS
::      ::      ::      ::      ::      ::      ::
HPRAALWAREP---HGQWEFMFREQRGDPINDPLAFRLSDARTLGLDLTTVMTERQSQLPEKYIGFYQIRKP
      540      550      560      570      580      590      600

      260      270      280      290      300      310
AIGLEDNASASPLLGIKRSWQT---RALLIKALIDQGYMKQILVSNOWLFGFSSYV-----TNIMDVMDRVNP
      ::      ::      ::      ::      ::      ::
PWLME-----QPPPPSRQTKPDAAATMPPPLSAGASVSALRYDDESWRPLSTVDDHKAWLDLDESHWVLG
      610      620      630      640      650      660

      320      330      340      X
D-GMAFIPLR-VIPFYERRASHRKRCQASL
      ::      ::      ::
DSRPDDIKQRRLLKATQRRGAEIDRFMPVVFEECYDQRF
      670      680      690      700
  
```

4. LOW344-FIG1.PEP

CYAA\$BORPE CALMODULIN-SENSITIVE ADENYLATE CYCLASE PRECURSOR (

```

ID  CYAA$BORPE      STANDARD;      PRT;  1706 AA.
AC  P15318;
DT  01-APR-1990 (REL. 14, CREATED)
DT  01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT  01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
DE  CALMODULIN-SENSITIVE ADENYLATE CYCLASE PRECURSOR (EC 4.6.1.1)
DE  (ATP PYROPHOSPHATE-LYASE) (ADENYLYL CYCLASE) (CYCLOLYSIN) (CONTAINS:
DE  HEMOLYSIN) (GENE NAME: CYA).
OS  BORDETELLA PERTUSSIS.
OC  PROKARYOTA; BACTERIA; GRAM-NEGATIVE AEROBIC RODS AND COCCI; UNCERTAIN.
RN  [1] (STRAIN 18323, SEQUENCE FROM N. A.)
RA  GLASER P., LADANT D., SEZER O., PICHOT F., ULLMANN A., DANCHIN A.;
RL  MOL. MICROBIOL. 2:19-30(1988).
CC  -!- FUNCTION: THIS ADENYLATE CYCLASE BELONGS TO A SPECIAL CLASS OF
CC  BACTERIAL TOXIN. IT ACTS ON MAMMALIAN CELLS BY ELEVATING CAMP-
CC  CONCENTRATION AND THUS DISRUPTS NORMAL CELL FUNCTION.
CC  -!- CATALYTIC ACTIVITY: ATP = 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC  -!- SUBCELLULAR LOCATION: RELEASED EXTRACELLULARLY IN A PROCESSED
CC  FORM.
CC  -!- DISEASE: WHOOPING COUGH.
DR  EMBL; Y00545; BPCYA.
DR  PIR; S00893; S00893.
  
```

DR PROSITE; PS00330; HEMOLYSIN_CALCIIUM.
 KW LYASE; CAMP SYNTHESIS; HEMOLYSIS; TOXIN; VIRULENCE; WHOOPING COUGH;
 KW CALCIUM-BINDING; REPEAT.
 FT CHAIN 1 1706 CALMODULIN-SENSITIVE ADENYLATE
 FT CYCLASE PRECURSOR.
 FT CHAIN 1 ? CALMODULIN-SENSITIVE ADENYLATE
 FT CYCLASE.
 FT CHAIN 1 1300 HEMOLYSIN, BY SIMILARITY TO E. COLI
 FT HEMOLYSINS (HYLA).
 FT DOMAIN 1 360 A, CALMODULIN-SENSITIVE CATALYTIC
 FT CENTRE.
 FT DOMAIN 361 912 B, ALA/GLY RICH.
 FT DOMAIN 913 1656 C.
 FT DOMAIN 1657 1706 D, ASP/GLY RICH.
 SQ SEQUENCE 1706 AA; 177505 MW; 1.25518E+07 CN;

Initial Score = 10 Optimized Score = 69 Significance = 4.34
 Residue Identity = 23% Matches = 93 Mismatches = 206
 Gaps = 97 Conservative Substitutions = 0

```

      X      10      20      30      40
      MQTRRVVLKSAAAGTLLGGLAGCATWLD RSAQAIGSIR-----ARPITI---SEA
      :      :      :      :      :      :      :      :      :
LMTQFGRAGSTNTPQEAASLSAAVFGL-GEASSAVAETVSGFFRGSSRWAGGFGVAGGAMALGGGIAAAVGA
      490      500      510      520      530      540      550

      50      60      70      80      90      100      110
      GFTLT HEDICG--SSAGFLRAWPEFFG-----SRKALAEKAVRGLRARAAGVRTIVDVSTFDIGRDVSL LAE
      :      :      :      :      :      :      :      :      :
      GMSLTDDAPAGQKAAAGAEIALQLTGGTVELASSIALA-----LAAARGVTSGLBVAGASAG-----
      560      570      580      590      600

      120      130      140      150      160      170
      VSRADVHIVAATGLWFDPPLSMRLRYVEELTQFFLREIQYGIE-----DTGIRAGI IKVATT
      :      :      :      :      :      :      :      :      :
      ---AAAGALAAALSPMEIYGLVQQSHYADQLDKLAGQESSAYGYEGDALLAQLYRDKTAAEGAVAGVSAVLST
      610      620      630      640      650      660      670

      180      190      200      210      220      230
      GKATPFQELVLKAAARASLATGVPVTTHTAASQRDG-----ERGRPPFLSPKLEPSRVCIGHSDDTDDL SYL
      :      :      :      :      :      :      :      :      :
      VQA-----AVSIAAA-AS-VVGAPVAVVT--SLLTGALNGILRGVQBP IIEKL-----ANDYARKIDELGGP
      680      690      700      710      720      730

      240      250      260      270      280      290      300
      TA-----LLRGYLIGLDHIPHSAIGLEDNASASPLLGI RSWQTRALLIKALIDQGYMKQILVSN DWLFGF
      :      :      :      :      :      :      :      :      :
      QAYFEKNLQARHEQLANS DGLRKMLADLBAGWNASSVIG---VQTEISK SAL-----ELAAITGNADNL--K
      740      750      760      770      780      790

      310      320      330      340      X
      SSVVTNIMDVM DRVNP DGM AFI P--LRVIPFYERRASHRKR CQASL
      :      :      :      :      :      :      :      :      :
      SVDV----FVDRFVQGERVAGQPVVLDVAAGGIDIAS-RKGERPALTFITPLAAPG
      800      810      820      830      840

```

5. LOW344-FIG1.PEP

CAP1\$MESCR PHOSPHOENOLPYRUVATE CARBOXYLASE 1 (EC 4.1.1.31) (G

ID CAP1\$MESCR STANDARD; PRT; 366 AA.
 AC P10490;
 DT 01-JUL-1989 (REL. 11, CREATED)
 DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
 DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
 DE PHOSPHOENOLPYRUVATE CARBOXYLASE 1 (EC 4.1.1.31) (GENE NAME: PPCA).

OS COMMON ICE PLANT (MESEMBRYANTHEMUM CRYSTALLINUM).
 OC EUKARYOTA; PLANTA; SPERMATOPHYTA; ANGIOSPERMAE.
 RN [1] (SEQUENCE FROM N. A.)
 RA RICKERS J., CUSHMAN J., MICHALOWSKI C., SCHMITT J., BOHNERT H. J.;
 RL MOL. GEN. GENET. 215:447-454(1989).
 RN [2] (SEQUENCE FROM N. A.)
 RA CUSHMAN J. C., BOHNERT H. J.;
 RL NUCLEIC ACIDS RES. 17:6745-6745(1989).
 CC -!- FUNCTION: TO FORM OXALOACETATE, A FOUR-CARBON DICARBOXYLIC ACID
 CC SOURCE FOR THE TRICARBOXYLIC ACID CYCLE.
 CC -!- CATALYTIC ACTIVITY: ORTHOPHOSPHATE + OXALOACETATE = H(2)O +
 CC PHOSPHOENOLPYRUVATE + CO(2).
 CC -!- PATHWAY: TRICARBOXYLIC ACID CYCLE.
 DR EMBL; X13660; MCPPCR.
 DR EMBL; X14587; MCPPCA.
 KW LYASE; CARBON DIOXIDE FIXATION; ALLOSTERIC ENZYME;
 KW TRICARBOXYLIC ACID CYCLE.
 SQ SEQUENCE 966 AA; 110659 MW; 4690045 CN;

Initial Score = 8 Optimized Score = 69 Significance = 4.34
 Residue Identity = 23% Matches = 94 Mismatches = 212
 Gaps = 90 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      MQTRRVVLKSAAAGTLLGGLAGCATWLDRSAQA---IGSIR---ARPITISEAGFTLTTHEDI
      |      |      |      |      |      |      |      |      |      |
SVRRSLLQKHGRIRDCLAQLYAKDITPDQKQELDEALQREIQAAFRTEIRRTQPTPQDEMRAQMSYFHETI
180      190      200      210      220      230      240

      60      70      80      90      100      110
CGSSAGFLRAWPEFFGSRKALAEKAVRGLRAR-----AAGVRTIVDVSTFDIGRDVSLLAEVSR
      |      |      |      |      |      |      |      |      |      |
WNGVPKFLR-----RLDTALK-NIGITERVPYNAPLIQFSSWMGGDRDGNPRVTPEVTRDVCLLA-RMM
250      260      270      280      290      300      310

      120      130      140      150      160      170
AADVHIVAATGLWFDPLSM-----RLR-YVEELTQFFLREIQYGIEDTGIRAGIIKVATTGKATPFQE--L
      |      |      |      |      |      |      |      |      |      |
AANMYFSQIDELMF--ELSMWRCTDELREAEELHKYSKRDSKHYIE-----FWKQIPSSPEPYR
      320      330      340      350      360

180      190      200      210      220      230
VLKA-----AARASLATGV---PV-TTHTAASQRDGERGRPPFLSP-KLEPSRVCIGHSDDTDDL
      |      |      |      |      |      |      |      |      |      |
VILADVDRDKLYYTRERSRQLLASEVSEIPVEATFTEIDQ-----FLEPLELCYRSLCAGDRPVADGS
370      380      390      400      410      420      430

      240      250      260      270      280      290      300
YL-----TALLRGYLIGLDHIPHSAIGLEDNASASPLLGIRSWQTRALLIKALIDQGYMKQILVSNWDLFGF
      |      |      |      |      |      |      |      |      |      |
LLDFMRQVATFGLCLVKLDIRQESERHTDVMDAITTHLGIGS--YRDWTEEKRD--WLLSELRGKRPLFGP
      440      450      460      470      480      490

      310      320      330      340      X
SSYVT-NIMDVMDRVNPDGMAFIPLRVIPFY-ERRASHRKRCQASL
      |      |      |      |      |      |      |      |      |
DLPRTEIADVLDLTIN--VIAELPSDSFGAYVISMATAPSDVLAVELLQRECKVKK
500      510      520      530      540 X      550

```

6. LOW344-FIG1.PEP
 MIS\$HUMAN MULLERIAN INHIBITING FACTOR PRECURSOR (MIS).

ID MIS\$HUMAN STANDARD; PRT; 560 AA.
 AC P03971;
 DT 23-OCT-1986 (REL. 02, CREATED)


```

      320      330      340      X
MDRVNPDGM-----AFIPLRVIPFY-----ER-RASHRKRCQASL
  |   |   |   |   |   |   |   |   |   |   |   |
NHVVLKLLKMBARGAALARPPCCVPTAYAGKLLISLSEERISAHVPMNVATECGCR
      510      520      530      540      550      X 560

```

7. LOW344-FIG1.PEP

ATXA\$LEIDO PROBABLE E1-E2 TYPE CATION ATPASE 1A (EC 3.6.1.-).

```

ID  ATXA$LEIDO      STANDARD;      PRT;      974 AA.
AC  P11718;
DT  01-OCT-1989 (REL. 12, CREATED)
DT  01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)
DT  01-OCT-1989 (REL. 12, LAST ANNOTATION UPDATE)
DE  PROBABLE E1-E2 TYPE CATION ATPASE 1A (EC 3.6.1.-).
OS  LEISHMANIA DONOVANI.
OC  EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA.
RN  [1] (SEQUENCE FROM N. A.)
RA  MEADE J. C. , SHAW J. , LEMASTER S. , GALLAGHER G. , STRINGER J. R. ;
RL  MOL. CELL. BIOL. 7:3937-3946(1987).
CC  -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
CC  -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC      (E1-E2 ATPASES).
CC  -!- SIMILARITY: THE TWO L. DONOVANI CATION-TRANSPORTING ATPASE GENES
CC      ARE 98% HOMOLOGOUS.
CC  -!- CAUTION: IN POSITION 351 THE N. A. SEQUENCE PREDICTS ARG, THE
CC      PROTEIN TRANSLATION SHOWN IN THREE PLACES IN THE PAPER GIVES
CC      LYS, WHICH IS CONSERVED IN ALL KNOWN E1-E2 ATPASES. WE HAVE
CC      USED LYS AFTER CONFIRMATION FROM THE AUTHORS.
DR  EMBL; M17889; LDCATP1.
DR  PROSITE; PS00154; ATPASE_E1_E2.
KW  HYDROLASE; ATP HYDROLYSIS; TRANSMEMBRANE; PHOSPHORYLATION;
KW  MAGNESIUM; ATP-BINDING.
FT  TRANSMEM      93      112      PUTATIVE.
FT  TRANSMEM      118     137      PUTATIVE.
FT  TRANSMEM      265     286      PUTATIVE.
FT  TRANSMEM      295     321      PUTATIVE.
FT  TRANSMEM      631     651      PUTATIVE.
FT  TRANSMEM      662     684      PUTATIVE.
FT  TRANSMEM      698     712      PUTATIVE.
FT  TRANSMEM      738     761      PUTATIVE.
FT  TRANSMEM      813     840      PUTATIVE.
FT  TRANSMEM      869     887      PUTATIVE.
FT  MOD_RES       351     351      PHOSPHORYLATION.
SQ  SEQUENCE      974 AA;  107448 MW;  5115862 CN;

```

```

Initial Score      =      18  Optimized Score =      68  Significance =   3.95
Residue Identity   =     24%  Matches          =      93  Mismatches   =   210
Gaps               =      82  Conservative Substitutions =      0

```

```

      X      10      20      30      40      50
MSTRR-----VVLKSAAAGTLLGGLAGCATWLDRSABAIGSIRARPITISEAGFTLTHE----
  |   |   |   |   |   |   |   |   |   |   |   |   |   |
FLDPPRPDTKDTIRRSKEYGVDVKMITGDHLLIAKEMC-RMLDLDPNILTADKLPQIKDANDLPEDLGEKYG
500      510      520      530      540      550      560

      60      70      80      90      100     110     120
DICGSSAGFLRAWPEFFGSRKALAEKAVRGLRAR---AAGVRTIVDVSTFDIGRDVSLLAEVSRADVHIVA
  |   |   |   |   |   |   |   |   |   |   |   |   |   |
DMMLSVGGFAQVFPE-----HKFMIVETLRGRGYTCAMTGDGVNDAPALKRADV---GIAVHGATDAARAA
570      580      590      600      610      620      630

      130     140     150     160     170
ATGLWFDPLSMRLRYVE-----ELTRFFLREIQVGIEDT---GIRAGIIKVATTGKA-----TPFQ---E

```


FT	CARBOHYD	982	562	POTENTIAL.
FT	CARBOHYD	994	934	POTENTIAL.
FT	CARBOHYD	1283	1289	POTENTIAL.
FT	CARBOHYD	1659	1659	POTENTIAL.
FT	CARBOHYD	2336	2336	POTENTIAL.
FT	CARBOHYD	2489	2489	POTENTIAL.
FT	CARBOHYD	2573	2573	POTENTIAL.
FT	CARBOHYD	2739	2739	POTENTIAL.
FT	CARBOHYD	2759	2759	POTENTIAL.
FT	CARBOHYD	2864	2864	POTENTIAL.
FT	CARBOHYD	2902	2902	POTENTIAL.
SB	SEQUENCE	3430	AA: 379640	MW: 2.098737E+07
				CN:

Initial Score = 10 Optimized Score = 68 Significance = 3.95
Residue Identity = 21% Matches = 87 Mismatches = 231
Gaps = 78 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      MQTRRVVLKSAAAGTLLGGLAGCAT-----WLDRSAQAIGSIRARPITISEAGFTLTHE-
      :      :      :      :      :      :      :      :      :      :
VESHGKIGATQAGRFSITPSAPSYTLKLGEYGEVTVDCPEPRSGIDTSAYYVMS-----VGEKSFLVHREW
440      450      460      470      480      490      500

      60      70      80      90      100
----DICGSSAG-----FLRAWPE---FFGSRKAL--AEKAVRGLRARAAGV----RTIVDVSTFDIGR
      :      :      :      :      :      :      :      :      :      :
FMDLNLPWSSAGSTTWRNRETLMFEFEPHATKQSVVALGSGEGALHQAALAGAIPEFSSNTVKLTSGHLKCR
510      520      530      540      550      560      570

      110      120      130      140      150      160      170
DVSLLAEVSRAADVHIVAATGLWFDPLSMRLRYVEELTQFFLREIQYGIEDTGIRAGIIKVATTGKATPFQ
      :      :      :      :      :      :      :      :      :      :
VKMEKLQKGGTTYGVCSKAFKFAFTPADTGHGTVVLEL-----QYTGTDGPCKVPISSVASLNDLTPVG
580      590      600      610      620      630

      180      190      200      210      220      230
ELVLKAAARASLATGVP-VTTHTAASGRDGERGRPPFLS-----PKLEPSRVCIGHSDDT-----
      :      :      :      :      :      :      :      :      :      :
RLV-----TVNPFVSVATANSKVLIELEPPFGDSYIVVGRGEGQINHHWHKSGSSIGKAFTTTLRGA
640      650      660      670      680      690      700

      240      250      260      270      280      290      300
DDLSYL--TALLRGYLIGLDHIPHSAIGLEDNASASPLLGISWQTRALLIKALIDQGYMKQILVSNWLFG
      :      :      :      :      :      :      :      :      :      :
QRLAALGDTAWDFGSGGVFTSVGKAIHQVFGGAFRSLFGGMSWITQGLL-GALLLWMGINARDRSIAMTFL
710      720      730      740      750      760      770

      310      320      330      340      X
FSSYVTNIMDMRVNPD-GMAFIPLRVIPFYERRASHRKRCQASL
      :      :      :      :      :
AVGGVLLFLSV--NVHADTGCAIDIGRQELRCGSGVFIHNDVEAWMDRYKFYPETP
780      790      800      810      X 820

```

9. LOW344-FIG1.PEP

ATXB\$LEIDO PROBABLE E1-E2 TYPE CATION ATPASE 1B (EC 3.6.1.-).

ID	ATXB\$LEIDO	STANDARD;	PRT;	974 AA.
AC	P12522;			
DT	01-OCT-1989	(REL. 12, CREATED)		
DT	01-OCT-1989	(REL. 12, LAST SEQUENCE UPDATE)		
DT	01-OCT-1989	(REL. 12, LAST ANNOTATION UPDATE)		
DE	PROBABLE E1-E2 TYPE CATION ATPASE 1B (EC 3.6.1.-).			
OS	LEISHMANIA DONOVANI.			
OC	EUKARYOTA; PROTOZOA; SARCOMASTIGOPHORA; MASTIGOPHORA.			
RN	[1] (SEQUENCE FROM N. A.)			

RA MEADE J. C., HUDSON K. M., STRINGER S. L., STRINGER J. R. ;
 RL MOL. BIOCHEM. PARASITOL. 33:81-92(1989).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + ORTHOPHOSPHATE.
 CC -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES).
 CC -!- SIMILARITY: THE TWO L. DONOVANI CATION-TRANSPORTING ATPASE GENES
 CC ARE 98% HOMOLOGOUS.
 DR EMBL; J04004; LDCATP2.
 DR PROSITE; PS00154; ATPASE_E1_E2.
 KW HYDROLASE; ATP HYDROLYSIS; TRANSMEMBRANE; PHOSPHORYLATION;
 KW MAGNESIUM; ATP-BINDING.
 FT TRANSMEM 93 112 PUTATIVE.
 FT TRANSMEM 118 137 PUTATIVE.
 FT TRANSMEM 265 286 PUTATIVE.
 FT TRANSMEM 295 321 PUTATIVE.
 FT TRANSMEM 631 651 PUTATIVE.
 FT TRANSMEM 662 684 PUTATIVE.
 FT TRANSMEM 698 712 PUTATIVE.
 FT TRANSMEM 738 761 PUTATIVE.
 FT TRANSMEM 813 840 PUTATIVE.
 FT TRANSMEM 869 887 PUTATIVE.
 FT MOD_RES 351 351 PHOSPHORYLATION.
 SQ SEQUENCE 974 AA; 107304 MW; 5132373 CN;

Initial Score = 18 Optimized Score = 68 Significance = 3.95
 Residue Identity = 24% Matches = 93 Mismatches = 210
 Gaps = 82 Conservative Substitutions = 0

```

      X           10           20           30           40           50
      MSTR-----VVLKSAAGTLLGGLAGCATWLDRSABAIGSIRARPITISEAGFTLTHE----
      ||         ||         ||         ||         ||         ||
    FLDPPRPDTKDTIRRSKEYGVVDVKMITGDHLLIAKEMC-RMLDLDPNILTADKLPQIKDANDLPEDLGEKYG
    500       510       520       530       540       550       560

      60           70           80           90           100          110          120
    DICGSSAGFLRAWPEFFGSRKALAEKAVRGLRAR---AAGVRTIVDVSTFDIGRDVSLLAEVSRADVHIVA
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
    DMMLSVGGFAQVFPE-----HKFMIVETLRQRGYTCAMTGDGVNDAPALKRADV--GIAVHGATDAARAA
    570       580       590       600       610       620       630

      130          140          150          160          170
    ATGLWFDPPLSMRLRYVE-----ELTQFFLREIQYGIEDT---GIRAGIIKVATTGKA-----TPFQ---E
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
    ADMVLTEPGLS---VVVEAMLVSREVFQRMLSFLTYRISATLQLVCFFFIACFSLTPKAYGSVDPNFQFFHL
    640       650       660       670       680       690       700

    180          190          200          210          220          230          240
    LVLKAAARASLATGVPVT---THTAASQRDGERGRPPFLSPKLEPSRVCIGHSDDTDDLSYLTALLRGYLIG
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
    PVLMFMLITLLNDGCLMTIGYDHVIPS---ERPQKWNL-PVVFVS-----ASILAAVACGSSLM
    710       720       730       740       750

    250          260          270          280          290          300          310
    LDHIPHSAIGLEDNASASPLIGIRSWQTR-----ALLIKALIDQGYMKQILVSNWDLFGFSS--YVTNIMDM
    ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
    L-----LWIGLE---GYSSQYVENSWFHRLGLAQLPQGLVVTMMYLK-ISIS-DFLTFLFSSRTGGHFFFYVP
    760       770       780       790       800       810

      320          330          340          X
    DRVNPDGMAFIPLRVIPFYERRAS--HKRRCQASL
    ||  ||  ||  ||  ||  ||  ||  ||
    PSPILFCGAIISLLV---STMAASFVHKSRPDNVLTEGLAWGQTN
    820       830       840       850       860
  
```

ID ATPO\$DENBI STANDARD; PRT; 511 AA.
AC P05492;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT 01-NOV-1988 (REL. 09, LAST ANNOTATION UPDATE)
DE ATP SYNTHASE ALPHA CHAIN, MITOCHONDRIAL (EC 3.6.1.34).
OS OENOTHERA BIENNIS.
OG MITOCHONDRION.
OC EUKARYOTA; PLANTA; SPERMATOPHYTA; ANGIOSPERMAE.
RN [1] (SEQUENCE FROM N. A.)
RA SCHUSTER W., BRENNICKE A.;
RL MOL. GEN. GENET. 204:29-35(1986).
CC -!- FUNCTION: THIS IS ONE OF THE 5 CHAINS OF THE ENZYMATIC COMPONENT
CC (COUPLING FACTOR CF(1)) OF THE MITOCHONDRIAL ATPASE COMPLEX.
DR EMBL; X04023; MIOBATPA.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA.
KW ATP SYNTHESIS; CF(1) COUPLING FACTOR; HYDROGEN ION TRANSPORT;
KW HYDROLASE; ATP-BINDING; MITOCHONDRION.
FT NP_BIND 171 178 ATP (BY SIMILARITY).
FT ACT_SITE 373 373 BY HOMOLOGY.
SQ SEQUENCE 511 AA; 55596 MW; 1250759 CN;

Initial Score = 9 Optimized Score = 68 Significance = 3.95
Residue Identity = 23% Matches = 91 Mismatches = 223
Gaps = 67 Conservative Substitutions = 0

```

      X 10      20      30      40      50      60
MQTRRVVLKSAAAGTLLGGLAGCATVLDRSAQ--AIG-SIR-----ARPITISEAGFTLTHTEDICGSSAGFL
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      MEFSPRAAELTTLESRTITNFYTNFQVDEIGRVISVGDGIARVYGLNEIQAGEMVEFASGVKGIAL
      X      10      20      30      40      50      60

      70      80      90      100     110     120
RAWPE-----FFGSRKALAEKAVRGRLRARAAGVRTIVDVSTFDIGRDVSLLAEVSRADVHI--VAATG----
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      NLENENVGIVVFGSDTAIKE---GDLVKR---TGSIVDV-----PAGKSLGRVVDALGVPIDGRGALGDHE
      70      80      90      100     110     120

      130     140     150     160     170     180     190
---LWFDPPLSMRLRYVEELTQFFL-REIQYGIEDTGIRAGIIKVATTGKATPFQELVLKAAARASLATGVP
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      RRRVEVKVPGIIERKSVHEPMQTGLKAVDSLVPIGRGQRELIIGDRQTGKTAIAIDTILNQKQMNSRATSES
      130     140     150     160     170     180     190

      200     210     220     230     240     250
VT---THTAASQ-RDGERGRPPFLS--PKLEPSRVCIGHSDDTDDLSYLTALLRGYLIG---LDHIPHSAI-
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      ETLYCVYVAIGQKRSTVAGLVQILSEGNALAYSILVAATASDPAPLQFL-APYSGCAMGEYFRDNGMHALII
      200     210     220     230     240     250     260     270

      260     270     280     290     300     310
--GLEDNASA---SPLLGIRSWQTRALLIKALIDQGYMKQIL----VSNDWLFGFSSYVTNIMDVMDRVNPD
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      YDDLKQAVAYRQMSLILRRFPGRCA----FPGDVFYLSRLLERAAKRSQDTGAGS--LTALPVIETQAGD
      280     290     300     310     320     330

      320     330     340     X
GMAFIPLRVIPFYERRASHRKRCQASL
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      VSAYIPTNVI-----SITDQGXCLETELFYRGIR
      340     350     X     360

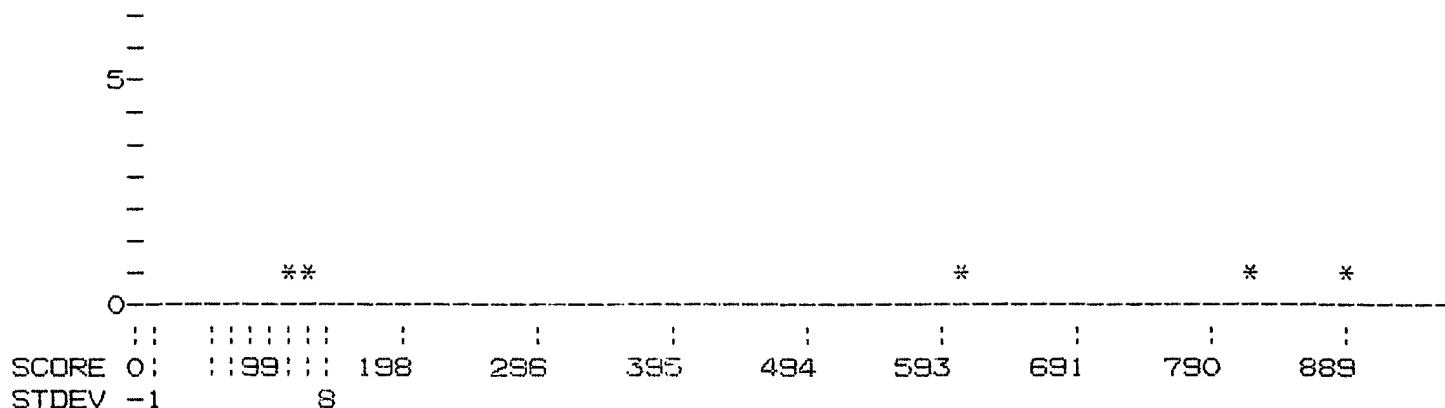
```

Results file low344-fig1.res made by maryh on Wed 17 Apr 91 12:41:07-PDT.

Query sequence being compared: LOW344-FIG1.SEQ
Number of sequences searched: 43406
Number of scores above cutoff: 156

Results of the initial comparison of LOW344-FIG1.SEQ with:
Data bank : GenBank 65.0, all entries
Data bank : UEMBL 24_65, all entries

100000-
--
N --
U50000-
M --
B --
E --
R --**
--
O --
F10000-
S --
E 5000- *
Q --
U --
E --
N --
C * *
E --
S 1000-
--
500-
-- *
--
--
--
100-
-- *
--
50-
--
--
--
--
10-
-- *



PARAMETERS

Similarity matrix	Unitary	K-tuple	4
Mismatch penalty	1	Joining penalty	30
Gap penalty	1.00	Window size	32
Gap size penalty	0.33		
Cutoff score	79		
Randomization group	0		
Initial scores to save	20	Alignments to save	10
Optimized scores to save	20	Display context	10

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	31	31	12.75

Times:	CPU	Total Elapsed
	00:54:30.02	02:41:00.00

Number of residues:	54775335
Number of sequences searched:	43406
Number of scores above cutoff:	156

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. PSEPTE	Plasmid pCMS1 (from P. diminuta)	1322	889	1304	67.29	0
2. M22863	Figure 1. Nucleotide sequence	1326	809	1309	61.02	0
3. FVBOPD	Flavobacterium sp. parathion h	1693	599	1281	44.55	0
4. X15898	Eimeria tenella mRNA for sporo	957	136	360	8.23	0
5. X14805	Mouse mRNA encoding DNA (cytos	4973	112	569	6.35	0
6. CELPOLII	C. elegans RNA polymerase II la	12993	111	526	6.27	0
7. HSHEPSH	Human hepatoma mRNA for serine	2363	108	562	6.04	0
8. HUMHPSNA	Human hepsin mRNA, complete cd	1783	108	456	6.04	0
9. BLYAMY2	Barley (H. vulgare) alpha-amyla	1588	107	590	5.96	0

10.	BLYAMYAA	Barley alpha-amylase type A is	1588	107	592	5.96	0
11.	BOVGABARB	Bovine mRNA for gamma-aminobut	3010	106	574	5.88	0
12.	BOVIGCAB	Bovine Ig germline gamma-2-cha	1979	104	479	5.73	0
13.	BTIGG2HC	Bovine Ig germline heavy chain	1979	104	479	5.73	0
14.	RRATP2	Rhodospirillum rubrum gene clu	4240	104	194	5.73	0
15.	PDEMDH	P. denitrificans methanol dehyd	2314	103	425	5.65	0
16.	MUSGT2A	M. musculus glucose transporter	2521	102	232	5.57	0
17.	SMASFUABC	S. marcescens periplasmic-bind	4583	101	583	5.49	0
18.	RABIGHAB	Rabbit Ig mu chain secreted fo	1953	101	489	5.49	0
19.	HUMIGHBD	Human Ig unproductively rearra	1127	100	384	5.41	0
20.	MUSRGEB3	Mouse 18S, 5.8S, 28S rRNA gene	3061	98	360	5.25	0

The scores below are sorted by optimized score.
Significance is calculated based on optimized score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
	**** 271 standard deviations above mean ****					
1. M22863	Figure 1. Nucleotide sequence	1326	809	1309	271.26	0
	**** 269 standard deviations above mean ****					
2. PSEPTE	Plasmid pCMS1 (from P. diminuta	1322	889	1304	269.45	0
	**** 261 standard deviations above mean ****					
3. FVBOPD	Flavobacterium sp. parathion h	1693	599	1281	261.10	0
	**** 15 standard deviations above mean ****					
4. TRN21TNPA	Transposon Tn21 tnpA gene for	3176	87	606	15.98	0
	**** 12 standard deviations above mean ****					
5. X17379	Sorghum vulgare mRNA for phosp	3147	80	597	12.71	0
	**** 11 standard deviations above mean ****					
6. NEUTRP1	n. crassa trifunctional tryptop	2750	85	594	11.62	0
	**** 10 standard deviations above mean ****					
7. BLYAMYAA	Barley alpha-amylase type A is	1588	107	592	10.89	0
8. PDUMER	Plasmid pDU1358 (from S. marces	2153	95	590	10.17	0
9. BLYAMY2	Barley (H. vulgare) alpha-amyla	1588	107	590	10.17	0
10. BPECYADE	Bordetella pertussis cyaD gene	2040	82	590	10.17	0
	**** 8 standard deviations above mean ****					
11. MUSNFMG	Mouse NF-M gene for middle-mol	5471	81	586	8.72	0
	**** 7 standard deviations above mean ****					
12. ACFTS140A	Fujinami sarcoma virus tempera	2715	86	584	7.99	0
13. SVGSII	Streptomyces viridochromogenes	2755	82	584	7.99	0
14. SMASFUABC	S. marcescens periplasmic-bind	4583	101	583	7.63	0
15. HS11UL	Herpes simplex virus type 1 (H	108360	94	582	7.26	0
	**** 6 standard deviations above mean ****					
16. ATUPRIREP	A. tumefaciens plasmid pRI4b r	4638	80	581	6.90	0
17. HUMASPX	Human nonerythroid alpha-spect	7787	80	581	6.90	0
18. SERCYSA	S. erythraea rhodanese-like pro	3373	81	581	6.90	0
19. X51950	E. coli purHD operon for AICAR	3432	94	580	6.54	0
20. MUSHCK	Mouse hck gene for tyrosine ki	1960	81	580	6.54	0

1. LOW344-FIG1.SEQ

M22863 Figure 1. Nucleotide sequence of Flavobacterium op

LOCUS M22863 1336 bp ds-DNA UNA 15-JUN-1989
 DEFINITION Figure 1. Nucleotide sequence of Flavobacterium opd gene fragment.
 ACCESSION M22863
 KEYWORDS .
 SOURCE
 ORGANISM Unknown
 Unclassified.
 REFERENCE 1 (bases 1 to 1326)

AUTHORS Harper, L. L. , McDaniel, C. S. , Miller, C. E. and Wild, J. R.
 TITLE Dissimilar Plasmids Isolated from *Pseudomonas diminuta* MG and a
Flavobacterium sp. (ATCC 27551) Contain Identical *opd* Genes
 JOURNAL Appl. Environ. Microbiol. 54, 2586-2589 (1988)
 STANDARD unannotated staff_entry
 BASE COUNT 279 a 363 c 392 g 286 t
 ORIGIN

Initial Score = 809 Optimized Score = 1309 Significance = 271.26
 Residue Identity = 98% Matches = 1316 Mismatches = 6
 Gaps = 7 Conservative Substitutions = 0

```

X      10      20      30      40      50      60      70
CTGCAGCCTGACTCGGCACCAAGTCTGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGCATGCAAACGA
|||||
CTGCAGCCTGACTCGGCACCAAGTCTGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGCATGCAAACGA
X      10      20      30      40      50      60      70

      80      90      100      110      120      130      140
GAAGGGTTGTGCTCAAGTCTGCGGCCGCGAGAACTCTGCTCGGCGGCCTGGCTGGGTGCGCGACGTGGCTGG
|||||
GAAGGGTTGTGCTCAAGTCTGCGGCCGCGAGAACTCTGCTCGGCGGCCTGGCTGGGTGCGCGACGTGGCTGG
      80      90      100      110      120      130      140

      150      160      170      180      190      200      210
ATCGATCGGCACAGGCGATCGGATCAATACGTGCGCGTCCTATCACAACTCTCTGAAGCGGGTTTCACACTGA
|||||
ATCGATCGGCACAGGCGATCGGATCAATACGTGCGCGTCCTATCACAACTCTCTGAAGCGGGTTTCACACTGA
      150      160      170      180      190      200      210

      220      230      240      250      260      270      280
CTCACGAGGACATCTGCGGCAGCTCGGCAGGATTCTTGCGTGCTTGGCCAGAGTTCTTCGGTAGCCGCAAG
|||||
CTCACGAGGACATCTCGGCAGCTCGGCAGGATTCTTGCGTGCTTGGCCAGAGTTCTTCGGTAGCGCAAG
      220      230      240      250      260      270      280

290      300      310      320      330      340      350      360
CTCTAGCGGAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGCTGGCGTGCGAACGATTGTGATGTGTGCGA
|||||
CTCTAGCGCAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGCTGGCGTGCGAACGATTGTGATGTGTGCGA
      290      300      310      320      330      340      350

      370      380      390      400      410      420      430
CTTTGATATCGGTGCGGACGTCAATTTATTGGCCGAGGTTTCGCGGGCTGCCGACGTTTCATATCGTGGCGG
|||||
CTTTGATATCGGTGCGGACGTCAATTTATTGGCCGAGGTTTCGCGGGCTGCCGACGTTTCATATCGTGGCGG
360      370      380      390      400      410      420

      440      450      460      470      480      490      500
CGACCGGCTTGTGGTTGACCCGCGCACTTTCGATGCGATTGAGGTATGTAGAGGAACTCACACAGTTCTTC
|||||
CGACCGGCTTGTGGTTGACCCGCGCACTTTCGATGCGATTGAGGTATGTAGAGGAACTCACACTAGTTCTTC
430      440      450      460      470      480      490      500

      510      520      530      540      550      560      570
CTGC-GTGAGATTCAATATGTCATCGAAG-ACACCGGAATTAGGGCGGGCATTATCAAGGTCGCGACCACAG
|||||
CTGCGGTGAGATCAATATGTCATCGAAGTACACCGGAATTAGGGCGGGCATTATCAAGGTCGCGACCACAG
      510      520      530      540      550      560      570

      580      590      600      610      620      630      640
GCAAGGCGACCCCTTTTCAGGAGTTAGTGTAAAGGCGGCCGCCCGGGCCAGCTTGGCCACCGGTGTTCCGG
|||||
GCAAGGCGACCCCTTTTCAGGAGTTAGTGTAAAGGCGGCCGCCCGGGCCAGCTTGGCCACCGGTGTTCCGG
      580      590      600      610      620      630      640

```

```

650      660      670      680      690      700      710
TAACCACTCACACGGCAGCAAGTCAGCGCGATGTTGAGCGAGGCAGGCCGCCATTTTGTAGTCCGAAGCTTG
|||||
TAACCACTCACACGGCAGCAAGTCAGCGCGATGTTGAGCGAGGCAGGCCGCCATTTTGTAGTCCGAAGCTTG
650      660      670      680      690      700      710

720      730      740      750      760      770      780
AGCCCTCACGGGTTTGATTTGGTCACACCGATGATACTGACGATTTGAGCTATCTCACCGCCCTGCTGCGCG
|||||
AGCCCTCACGGGTTTGATTTGGTCACACCGATGATACTGACGATTTGAGCTATCTCACCGCCCTGCTGCGCG
720      730      740      750      760      770      780

790      800      810      820      830      840      850      860
GATACCTCATCGGTCTAGACCACATCCCGCACAGTGCGATTGGTCTAGAAGATAATGCGAGTGCATCACCGC
|||||
GATACCTCATCGGTCTAGACCACATCCCGCACAGTGCGATTGGTCTAGAAGATAATGCGAGTGCATCACCGC
790      800      810      820      830      840      850      860

870      880      890      900      910      920      930
TCCTGGGCATCCGTTTGGTGGCAAACACGGGCTCTCTTGATCAAGGCGCTCATCGACCAAGGCTACATGAAAC
|||||
TCCTGGGCATCCGTTTGGTGGCAAACACGGGCTCTCTTGATCAAGGCGCTCATCGACCAAGGCTACATGAAAC
870      880      890      900      910      920      930

940      950      960      970      980      990      1000
AAATCCTCGTTTTCGAATGACTGGCTGTTCGGGTTTTTCGAGCTATGTCACCAACATCATGGACGTGATGGATC
|||||
AAATCCTCGTTTTCGAATGACTGGCTGTTCGGGTTTTTCGAGCTATGTCACCAACATCATGGACGTGATGGATC
940      950      960      970      980      990      1000

1010      1020      1030      1040      1050      1060      1070
GCGTGAACCCCGACGGGATGGCCTTCATTCCACTGAGAGTGATCCCATTTCTACGAGAGAAGGGCGTCCCACA
|||||
GCGTGAACCCCGACGGGATGGCCTTCATT-CACTGAGAGTGATCCCATTTCTACGAGAGAAGGGCGTCCCACA
1010      1020      1030      1040      1050      1060      1070

1080      1090      1100      1110      1120      1130      1140
GGAAACGCTGCCAGGCATCACTGTGACTAACCCGGCGCGGTTTCTGTGTACCGACTTGCCGTGCATGACGCC
|||||
GGAAACGCTGCCAGGCATCACTGTGACTAACCCGGCGCGGTTTCTGTGTACCGACTTGCCGTGCATGACGCC
1080      1090      1100      1110      1120      1130      1140

1150      1160      1170      1180      1190      1200      1210      1220
ATCTGGATCCTTCCACGCAGCGGCCACTATTCCCCGTCAAGATACCGAACGATGAAGTCGCGCATCGATCGA
|||||
ATCTGGATCCTTCCACGCAGCGGCCACTATTCCCCGTCAAGATACCGAACGATGAAGTCGCGCATCGATCGA
1150      1160      1170      1180      1190      1200      1210      1220

1230      1240      1250      1260      1270      1280      1290
TAGGCATCTTCAATGTGATCAGGGCTGCCACCTCCAAAGCCGGTGGCCACCCCTGTTCGATAGTCTTGAGGGA
|||||
TAGGCATCTTCAATGTGATCAGGGCTGCCACCTCCAAAGCCGGTGGCCACCCCTGTTCGATAGTCTTGAGGGA
1230      1240      1250      1260      1270      1280      1290

1300      1310      1320      X
CGGTAGCGACGACCGTGCTTTTCGTGAACTGCAG
|||||
CGGTAGCGACGACCGTGCTTTTCGTGAACTGCAG
1300      1310      1320      X

```

2. LOW344-FIG1.SEQ

PSEPTE Plasmid pCMS1 (from *P. diminuta*) phosphodiesterase

LOCUS PSEPTE 1322 bp ds-DNA BCT 15-MAR-1989
 DEFINITION Plasmid pCMS1 (from *P. diminuta*) phosphodiesterase (opd) gene,

```

complete cds.
ACCESSION M20392
KEYWORDS phosphotriesterase.
SOURCE Plasmid pCMS1 (from Pseudomonas diminuta) DNA.
ORGANISM Plasmid pCMS1
          Prokaryota; Bacteria; Plasmid pCMS1.
REFERENCE 1 (bases 1 to 1322)
AUTHORS McDaniel, C. S., Harper, L. L. and Wild, J. R.
TITLE Cloning and sequencing of a plasmid-borne gene (opd) encoding a
       phosphotriesterase
JOURNAL J. Bacteriol. 170, 2306-2311 (1988)
STANDARD simple staff_review
FEATURES             Location/Qualifiers
     CDS             63..1040
                     /note="phosphotriesterase protein"
BASE COUNT      278 a      367 c      392 g      285 t
ORIGIN           5 bp upstream from PstI site.

```

```

Initial Score      =      889   Optimized Score =      1304   Significance = 269.45
Residue Identity  =      98%   Matches          =      1313   Mismatches   =        6
Gaps              =         11   Conservative Substitutions =          0

```

```

X          10          20          30          40          50          60          70
CTGCAGCCTGACTCGGCACCAAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGCATGCAAACGA
|||||
CTGCAGCCTGACTCGGCACCAAGTCGCTGCAAGCAGAGTCGTAAGCAATCGCAAGGGGGCAGCATGCAAACGA
X          10          20          30          40          50          60          70

          80          90         100         110         120         130         140
GAAGGGTTGTGCTCAAGTCTGCGGCCGCGAGGAAGTCTGCTGGGCGGCCTGGCTGGGTGCGCGACGTGGCTGG
|||||
GAAGGGTTGTGCTCAAGTCTGCGGCCGCGAGGAAGTCTGCTGGGCGGCCTGGCTGGGTGCGCGACGTGGCTGG
          80          90         100         110         120         130         140

          150         160         170         180         190         200         210
ATCGATCGGCACAGGCGATCGGATCAATACGTGCGCGTCCTATCACAATCTCTGAAGCGGGTTTCACACTGA
|||||
ATCGATCGGCACAGGCGATCGGATCAATACGTGCGCGTCCTATCACAATCTCTGAAGCGGGTTTCACACTGA
          150         160         170         180         190         200         210

          220         230         240         250         260         270         280
CTCACGAGGACATCTGCGGCAGCTCGGCAGGATTCTTGCGTGCTTGCCAGAGTTCTTCGGTAGCCGCAAAG
|||||
CTCACGAGGACATCTCGGCAGCTCGGCAGGATTCTTGCGTGCTTGCCAGAGTTCTTCGGTAGCGCAAAG
          220         230         240         250         260         270         280

290         300         310         320         330         340         350         360
CTCTAGCGGAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGCTGGCGTGCGAACGATTGTTCGATGTGTCTGA
|||||
CTCTAGCGGAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGCTGGCGTGCGAACGATTGTTCGATGTGTCTGA
          290         300         310         320         330         340         350

          370         380         390         400         410         420         430
CTTTGATATCGGTGCGGACGTCAGTTTATTGGCCGAGGTTTCGCGGGCTGCCGACGTTTCATATCGTGGCGG
|||||
CTTTGATATCGGTGCGGACGTCAGTTTATTGGCCGAGGTTTCGCGGGCTGCCGACGTTTCATATCGTGGCGG
          360         370         380         390         400         410         420

          440         450         460         470         480         490         500
CGACCGGCTTGTGGTTCGACCCGCCACTTTTCGATGCGATTGAGGTATGTAGAGGAACTCACACAGTTCTTC
|||||
CGACCGGCTTGTGGTTCGACCCGCCACTTTTCGATGCGATTGAGGTATGTAGAGGAACTCACACTAGTTCTTC
          430         440         450         460         470         480         490         500

          510         520         530         540         550         560         570
CTGC-GTGAGATTCAATATGGCATCGAAG-ACACCGGAATTAGGGCGGGCATTATCAAGGTGCGGACCACAG

```

```

|||||
CTGCGGTGAGATTCAATATGTCATCGAAGTACACCGGAATTAGGGCGGGCATTATCAAGGTCGCGACCACAG
  510      520      530      540      550      560      570

      580      590      600      610      620      630      640
GCAAGGCGACCCCTTTTCAGGAGTTAGTGTTAAGGCGGGCCCGGGCCAGCTTGGCCACCGGTGTTCCGG
|||||
GCAAGGCGACCCCTTTTCAGGAGTTAGTGTTAAGGCGGGCCCGGGCCAGCTTGGCCACCGGTGTTCCGG
  580      590      600      610      620      630      640

      650      660      670      680      690      700      710
TAACCACTCACACGGCAGCAAGTCAGCGCGATGGTGAGCGAGGCGAGGCCGCCATTTTGTAGTCCGAAGCTTG
|||||
TAACCACTCACACGGCAGCAAGTCAGCGCGATGGTGAGCGAGGCGAGGCCGCCATTTTGTAGTCCGAAGCTTG
  650      660      670      680      690      700      710

      720      730      740      750      760      770      780
AGCCCTCACGGGTTTGTATTGGTCACACCGATGATACTGACGATTTGAGCTATCTCACCGCCCTGCTGCGCG
|||||
AGCCCTCACGGGTTTGTATTGGTCACACCGATGATACTGACGATTTGAGCTATCTCACCGCCCTGCTGCGCG
  720      730      740      750      760      770      780

790      800      810      820      830      840      850      860
GATACCTCATCGGTCTAGACCACATCCCGCACAGTGCGATTGGTCTAGAAGATAATGCGAGTGCATCACCGC
|||||
GATACCTCATCGGTCTAGACCACATCCCGCACAGTGCGATTGGTCTAGAAGATAATGCGAGTGCATCACCGC
790      800      810      820      830      840      850      860

      870      880      890      900      910      920      930
TCCTGGGCATCCGTTTCTGTGGCAAACACGGGCTCTCTTGATCAAGGCGCTCATCGACCAAGGCTACATGAAAC
|||||
TCCTGGGCATCCGTTTCTGTGGCAAACACGGGCTCTCTTGATCAAGGCGCTCATCGACCAAGGCTACATGAAAC
  870      880      890      900      910      920      930

      940      950      960      970      980      990      1000
AAATCCTCGTTTTCGAATGACTGGCTGTTCGGGTTTTCGAGCTATGTCACCAACATCATGGACGTGATGGATC
|||||
AAATCCTCGTTTTCGAATGACTGGCTGTTCGGGTTTTCGAGCTATGTCACCAACATCATGGACGTGATGGATC
  940      950      960      970      980      990      1000

      1010      1020      1030      1040      1050      1060      1070
GCGTGAACCCCGACGGGATGGCCTTCATTCCACTGAGAGTGATCCCATTTCTACGAGAGAAGGGCGTCCCACA
|||||
GCGTGAACCCCGACGGGATGGCCTTCATTCCACTGAGAGTGATCCCATTTCTACGAGAGAAGGGCGTCCCACA
  1010      1020      1030      1040      1050      1060      1070

1080      1090      1100      1110      1120      1130      1140
GGAAACGCTGCCAGGCATCACTGTGACTAAGCCGCGCGGGTTCTGTGTACCGACTTGCCGTGCATGACGCC
|||||
GGAAACGCTGCCAGGCATCACTGTGACTAAGCCGCGCGGGTTCTGTGTACCGACTTGCCGTGCATGACGCC
  1080      1090      1100      1110      1120      1130      1140

1150      1160      1170      1180      1190      1200      1210      1220
ATCTGGATCCTTCCACGCGAGCGGCCACTATTCCCGGTCAAGATACCGAACGATGAAGTCGCGCATCGATCGA
|||||
ATCTGGATCCTTCCACGCGAGCGGCCACTATTCCCGGTCAAGATACCGAACGATGAAGTCGCGC-----ATCGA
  1150      1160      1170      1180      1190      1200      1210

      1230      1240      1250      1260      1270      1280      1290
TAGGCATCTTCAATGTGATCAGGGGTGCGAAGCTCCAAAGCGGGTGGCCACCCCTGTGATAGTCTTGAGGGA
|||||
TAGGCATCTTCAATGTGATCAGGGGTGCGAAGCTCCAAAGCGGGTGGCCACCCCTGTGATAGTCTTGAGGGA
  1220      1230      1240      1250      1260      1270      1280

      1300      1310      1320      %
CGGTAGCGACGACCGTGGCTTTTGGTGAAGCTGAG

```


640 650 660 670 680 690 700
 350 360 370 380 390 400 410
 ACGATTGTCGATGTGTGCGACTTTTCGATATCGGTGCGGACGTCAGTTTATTGGCCGAGGTTTCGCGGGCTGCC

 ACGATTGTCGATGTGTGCGACTTTTCGATATCGGTGCGGACGTCAGTTTATTGGCCGAGGTTTCGCGGGCTGCC
 710 720 730 740 750 760 770
 420 430 440 450 460 470 480
 GACGTTTCATATCGTGGCGGCGACCGGCTTGTGATTTCGACCCGCCACTTTTCGATGCGATTGAGGTATGTAGAG

 GACGTTTCATATCGTGGCGGCGACCGGCTTGTGATTTCGACCCGCCACTTTTCGATGCGATTGAGGTATGTAGAG
 780 790 800 810 820 830 840 850
 490 500 510 520 530 540 550
 GAACTCACACAGTTCTTCCTGCGTGAGATTCAATATGGCATCGAAGACACCGGAATTAGGGCGGGCATTATC

 GAACTCACACAGTTCTTCCTGCGTGAGATTCAATATGGCATCGAAGACACCGGAATTAGGGCGGGCATTATC
 860 870 880 890 900 910 920
 560 570 580 590 600 610 620
 AAGGTCGCGACACAGGCAAGGCGACCCCTTTTCAGGAGTTAGTGTTAAAGGCGGCCGCCCGGGCCAGCTTG

 AAGGTCGCGACACAGGCAAGGCGACCCCTTTTCAGGAGTTAGTGTTAAAGGCGGCCGCCCGGGCCAGCTTG
 930 940 950 960 970 980 990
 630 640 650 660 670 680 690 700
 GCCACCGGTGTTCCGGTAACCACTCACACGCGCAGCAAGTCAGCGCGATGGTGAGCGAGGCAGGCCGCCATTT

 GCCACCGGTGTTCCGGTAACCACTCACACGCGCAGCAAGTCAGCGCGATGGTGAGC-AG-CAGGCCGCCATTT
 1000 1010 1020 1030 1040 1050 1060
 710 720 730 740 750 760 770
 TTGAGTCCGAA-GCTTGAG-CCCTCAGCGGTTTGTATTGGTCACAGCGATGATACTGACGATTTGAGCTATC

 TTGAGTCCGAAGGCTTGAGCCCTCAGCGGTTTGTATTGGTCACAGCGATGATACTGACGATTTGAGCTATC
 1070 1080 1090 1100 1110 1120 1130
 780 790 800 810 820 830 840
 TCACCGCCCT-GCT--GCGCGGATACCTCATCGGTCTAGACCACATCCCGCACAGTGCGATTGGTCTAGAAG

 TCACCGCCCTCGCTGCGCGGATACCTCATCGGTCTAGACCACATCCCGCACAGTGCGATTGGTCTAGAAG
 1140 1150 1160 1170 1180 1190 1200
 850 860 870 880 890 900 910
 ATAATGCGAGTGCATCACCGCTCCTGGGCATCCGTTTCGTGGCAACACGGGCTCTCTTGATCAAGGCGCTCA

 ATAATGCGAGTGCATCACCGCTCCTGGGCATCCGTTTCGTGGCAACACGGGCTCTCTTGATCAAGGCGCTCA
 1210 1220 1230 1240 1250 1260 1270 1280
 920 930 940 950 960 970 980
 TCGACCAAGGCTACATGAACAAATCCTCGTTTCGAATGACTGGCTGTTTCGGGTTTTTCGAGCTATGTCACCA

 TCGACCAAGGCTACATGAACAAATCCTCGTTTCGAATGACTGGCTGTTTCGGGTTTTTCGAGCTATGTCACCA
 1290 1300 1310 1320 1330 1340 1350
 990 1000 1010 1020 1030 1040 1050
 ACATCATGGACGTGATGGATCGCGTGAACCCCGACCGGATGGCCTTCATTCCACTGAGAGTGATCCCATT-C

 ACATCATGGACGTGATGGATCGCGTGAACCCCGACCGGATGGCCTTCATTCCACTGAGAGTGATCCCATTCC
 1360 1370 1380 1390 1400 1410 1420
 1060 1070 1080 1090 1100 1110 1120
 TACGAGAGAAGGGCGTCCCACAGGAAACGCTGCCAGGCATCACTGTGACTAACCCGGCGCGGTTCTGTGTCA

 TACGAGAGAAGGGCGTCCCACAGGAAACGCTGCCAGGCATCACTGTGACTAACCCGGCGCGGTTCT-TGTCA

```

1430      1440      1450      1460      1470      1480      1490
1130      1140      1150      1160      1170      1180      1190
CCGA-CTTGC---CGTGCATGACGCCATCTGGATCCTTCCAGCGAGCGGCCACTATTCCCCGTCAAGATACG
||||| ||||| ||| ||||||||||||||||||||||||||| |||||||||||||||||||||||||||
CCGACCTTGCGGGCGT-CATGACGCCATCTGGATCCTTCCAGCGAGCGGCCACTATTCCCCGTCAAGATACC
1500      1510      1520      1530      1540      1550      1560

1200      1210      1220      1230      1240      1250      1260
GAACGATGAAGTCGCGCATCGATCGATAGGCATCTTCAATGTGATCAGGGCTGCCACCTCCAAAGCCGGTGG
||||| ||||||||||||||||||||||||||| ||||||||||||||||||||||||||| ||||
GAACGATGAAGTCGCGCATCGATCGATAGGCATCTTCAATTTGATCAGGGCTGCCACCTCCAAAGCC-GTGG
1570      1580      1590      1600      1610      1620      1630

1270      1280      1290      1300      1310      1320      X
CCACCCCTGTCGATAGTCTTGAGGGACGGTAGCGACGACCGTGCTTTTTCGTGAACTGCAG
||||| |||||||||||||| ||| ||| | |||||||||| |||||||
CCACCCCTGTCGATAGTCTTGA-GGAC-GTAGGGCACACCGTGCTTTTTC--GAACTGCAG
1640      1650      1660      1670      1680      1690      X

```

4. LOW344-FIG1. SEQ

TRN21TNPA Transposon Tn21 tnpA gene for transposase.

```

LOCUS       TRN21TNPA                3176 bp ds-DNA            BCT             15-MAR-1988
DEFINITION  Transposon Tn21 tnpA gene for transposase.
ACCESSION   X04891
KEYWORDS    tnpA gene; tnpR gene; transposase; transposon.
SOURCE      Transposon Tn21.
  ORGANISM  Transposon Tn21
            Prokaryota; Bacteria; Transposon Tn21.
REFERENCE   1 (bases 1 to 3176)
  AUTHORS   Ward,E. and Grinstead,J.
  TITLE     The nucleotide sequence of the tnpA gene of Tn21
  JOURNAL   Nucleic Acids Res. 15, 1799-1806 (1987)
  STANDARD  simple automatic
COMMENT     [1] enum. 1 to 3176.
FEATURES             Location/Qualifiers
     CDS             1..54
                     /note="tnpR gene"
     CDS             57..3023
                     /note="transposase (AA 1-988)"
BASE COUNT    652 a    1057 c    915 g    552 t
ORIGIN

```

```
Initial Score      =      87  Optimized Score      =     606  Significance      =    15.98
Residue Identity  =     52%  Matches              =     761  Mismatches        =     472
Gaps              =     212  Conservative Substitutions =         0
```

X 10 20 30 40 50
CTGCAGCCTGACTCGGCA--CCAG-TCGC--TGCAAG---CAGAGTCGTAAAGCAATCGCAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
TCCATCCTGTCCGCCGCC-GAGCGGGAAAGCCTGCTGGCGTTGCCGGACTCCAAGGACG-ACCTGATC-CGA
70 80 90 100 110 120 130

60 70 80 90 100 110
GGGGGCAGCATGCAAACGAGA-AGGGTTGT-GCTCA--AGTCTGCGGC-CG---CAGGAACTCTGCTGGGCG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
CATTACA-CATTC-AACGATACCGACCTCTCGATCATCCGACAGCGGCGCGGGCCAGCCAATCGGCTGGGCT
140 150 160 170 180 190 200

120 130 140 150 160 170 180
GCCTGGCTG--GATGCGCGACGTG-GCTGGATCGATCGGCA--CAGGCGATCGGATCAA-TA-CGTGCGCGT
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
TCGCGG-TGCAGCTCTGTTACCTGCGCTTTCCCG-ECGTATCCTGGGCGTC-GATGAACTACCGTTTTCCCG
210 220 230 240 250 260 270

[illegible]

ORIGIN

Initial Score = 80 Optimized Score = 597 Significance = 12.71
Residue Identity = 51% Matches = 740 Mismatches = 490
Gaps = 203 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      CTGC-AGC-CTGACT-CGCGACCCAGTCTGCT--GCAAG-CAGAGTCGTAAGCAATCG-CAAGG
      || ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
1460 1470 1480 1490 1500 1510 1520 1530
      GTCGGGTGAAAGGCAAGCGCCCACTGCTGCCCGCGACCTTCCCATGACCGAGGAGATCGCCGACGTCATCG
      || ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
      GCGC--CATGCGCGTCTCTGGCCGAGCTCCCGATCGAGAGCTTCGGCCCTACATCATCTCCATG--TGCACG
      || ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
1540 1550 1560 1570 1580 1590
      GCTGGGTGCGCGACGTGGCTGGATCGATCG-CCACAGGCGATCGGATCAATACGTGCGCGTCTCTATCACAA
      || ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
      GC-GCCCTCG-GACGT-GCTCG-CCG-TCGAGCTCCTGC-AGCG---CGAGATGTG-GCATTGCGCCAGCGGT
1600 1610 1620 1630 1640 1650 1660
      CTC--TGAAGCGGGTTTCACACTGACTCACGAGGACATCTGC-GGCAGCTCGGCAGGATTCTTG----CGTG
      || ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
      CCGCGTGGTGCCGCTGTTTCA--GAGGCTGGCCGAC--CTGCAGGCGGCGCGCGCG--TCCGTGGAGAAGCT
1670 1680 1690 1700 1710 1720
      CTTGGCCAGAGTTCTTCGGTAGCCGCAA--AGCTCTAGCGGAAAAGGCTGTGAGAGGATTGCGCGCCAGAGC
      || ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
      CTTCTCCACTG-ACT--GGTA-CATCAACCACATC-AACGGCAA--GCAG-CAGGTGATGGTCTGGCTACTCC
1730 1740 1750 1760 1770 1780 1790
      GGCTGGCGTGCGAACGATTGTCTG-ATGT-GTCGACTTTCGATATCGGT-CGCGACGTCAGTTTA-TTGGCCG
      || ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
      GACTCCGGCAAGGACGCGCGCCGCGCTGTCCGCGGCGT--GGCAGCTGTACGTGGCGCAGGAGGAGATGGCCA
1800 1810 1820 1830 1840 1850 1860
      AGGTTTCGCGGGCTGCCGACGTTTCATATCGTGGCGGCGACCGGCTTGTGGTTTCGACC-CG---CCACTTTCG
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      AGGT-----GGC-CAAGAAGT--ACGCGGTGAAGGTGAC---CTTGTTCCACGGCCGCGGTGGCACCGTCTG
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1870 1880 1890 1900 1910 1920
      ATGCGATTGAGGTATGTAGAGGAAGTACACAGTTCTT--CCTG-CGTGAGATTCAAT--ATGG---CATCG
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      --GCAGGGGGCGTTGGCCCGACGCAC-CTCGCCATCCGTGCCAGCGCGCGGACACCATCAACGGGTCCATCC
1930 1940 1950 1960 1970 1980 1990
      AAGACACCGGAATTAGGGCGGGCATTATCAAGGTGCGGACCACAGGCAAGGCGACCC---CCTTTCAGGAGT
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      TGGTGAC--GGTGCAGGGCGAG-GTCATCGAETTCATGTTT---GGGGAGGAGAACCTGTGCTTCCAGTCTC
2000 2010 2020 2030 2040 2050
      T--AG-TGTTAAAGGCGGCGCGC-CCGGGCGA-GC-TTG--GCCACCGGTGT-TCCGGTAACCACTCACACGG
      || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      TGCAGCGGTTTACGGGCGCCACGCTGGAGGACGCCATGCACCCGCGCGTCTCTCCCG-AGTGGCGCAATATG
2060 2070 2080 2090 2100 2110 2120
```


6. LOW344-FIG1. SEQ

NEUTRP1 n.crassa trifunctional tryptophan biosynthesis gen

LOCUS NEUTRP1 2750 bp DNA PLN 01-AUG-1983
 DEFINITION n.crassa trifunctional tryptophan biosynthesis gene trp-1.
 ACCESSION J01252
 KEYWORDS multifunctional enzyme.
 SOURCE neurospora crassa.
 ORGANISM Neurospora crassa
 Eukaryota; Plantae; Thalloblonta; Eumycota; Ascomycotina;
 Pyrenomycetes; Sordariales; Sordariaceae; Neurospora; crassa.
 REFERENCE 1 (bases 1 to 2750)
 AUTHORS Schechtman, M. G. and Yanofsky, C.
 TITLE structure of the trifunctional trp-1 gene from neurospora crassa
 and its aberrant expression in escherichia coli
 JOURNAL J. Mol. Appl. Genet. 2, 83-99 (1983)
 STANDARD simple staff_review
 BASE COUNT 572 a 821 c 727 g 630 t
 ORIGIN

Initial Score = 85 Optimized Score = 594 Significance = 11.62
 Residue Identity = 51% Matches = 738 Mismatches = 487
 Gaps = 213 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      CTGCAGCCTGACTCGGC-ACCAGTCGCTGCAAGCAGAGTCGTAAGCAATCG--CAAGGGGGC
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ATCACAATGTCGTCCTCCTCAGTCGTCGACCACTCTCCCA---CGATTTCGCTCCTTCGCCCTGGTGCC
  260      270      280      290      300      310      320

60      70      80      90      100      110
A---GCATGCAAACGAGAAAGGTTGTGCTCAAGTCTG---CG-GCCGCAGGAAC-TCTGCTGG---GCGGC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AACCGCCTCCAACC--TCATCCTCATCGACAACATGATTGCTTTACCTGGAACGTCTACCAGTACCTCGTC
  330      340      350      360      370      380      390

120      130      140      150      160      170      180
CTGGCTGGGTGCGCGACGTGGCTGGATCGATCGGCACAGGCGATCGGATCA--AT--ACGTGCGCGTCCTAT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CTCG-AGGGCGC-CAAGGTGACCG--TC-TTCCGCA---ACGACCACATCACCATCGACGAGCTCATCGCA-
  400      410      420      430      440      450

190      200      210      220      230      240      250
CACAAATCTC-TGAAGCGGGT-TTCA-CACTGACTCACGAGGACATCTGCGGCAGCTCGGCAGGATTCTTGCG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AAGAACCCACCCAGCTCGTCATCAGCCCTG-GGCCCG--GTCATC-CCGGCACCAGACTCCGGTATCTCGCG
  460      470      480      490      500      510      520

260      270      280      290      300      310      320
TGCTTGGC--CAGAGTTCTTCGGTAGCCGCAAGCTCTAGCGGAAAAGGCTGTGAGAGGATTGCGCGCCAGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CG-ATGCCATCAG-GCACTTC---GCCGGCAAGATC-CCCATCTTTGGC-GT--GTGCATGGGCCAGCAGT
  530      540      550      560      570      580

330      340      350      360      370      380      390
GCGGC--TGGCGTGCGAACGATTGTGATGTGTGACTTTGATATCGGTGCGGACGTGAGTTTATTGGCCG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCATCTTTGACGT-CTATGG--CGGCGACGTGT--GCT-TCG---CCGGT-GAGA-TTC--TGCACGGAAAG
  590      600      610      620      630      640

400      410      420      430      440      450      460
AGGTTTTCGCGGGCTGCCGACGTTTCATATCGTGCGGCGACCGGCTTGTGGTTTCGACCCGCCACTTTC--GAT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACCTCTCCTCTGC-GCC-ACGACCGCAAGCGCGCATATGCCGCTCTGTCTCAGGATCTGCCAGTGACGAGAT
  650      660      670      680      690      700      710

```

470 480 490 500 510 520
GCGATTGAGGTATGTAGAGGAACCTCA---CACAGTTCTTCC---TGCGT-GAGATTCAATATGGCAT--CGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
ACCACT---CTCT-TGCCGGTACTCATEITCACCCCTTC--CCGAGTGCTTGAGGTTACCTCTTGATTGCGA
720 730 740 750 760 770 780

530 540 550 560 570 580 590
A-GACACCGGAATTAGGGCGGGCATTATCAAGGTCGCGACCACAGGCAAGGCG-ACCCCCCTTTCAGGAGTT-
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGGAGGACGGTTCCAAGGGTGTGCATCATGGGTGTC-CG-CCA---CAAGGAGTACACCATTGAGGGTGTTC
790 800 810 820 830 840

600 610 620 630 640 650 660
AGTGTTAAAGGCGGCCGCCGGGCCAGCT--TGGCCACCG-GTGTTCGGTAAC--CACTCACACGGCAGCA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGTTCACCCGGAGAGTATTCTGTCTGCTGAGGGTCTGTGGCATGTTCCGG-AACTTCCTTCACA-TGCAGGG
850 860 870 880 890 900 910

670 680 690 700 710 720
AGTCAGCGCGATGGTGAGCGAG-----GC---AGGCCGCCATTTTTGAGTC-----CGAAGCTTGA-GC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGGCACTTGGGCGGAGAACGAGAGACTGCAAAGGCCGCCAGGCACAGGCTGCCAACACAAAGTCCGACGC
920 930 940 950 960 970 980 990

730 740 750 760 770 780 790
CCTCACGGGTTTGTATTGGTCACAGCGATGATACTGACGATTTGAGCTATCTCACC GCCCTGCTGC-GCGGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
TCCCACGCCCAAGAAGAG----CAAC-ATCCTTCAAAGATTACG----CCCACCGTAAGGCTGCTGTGGA
1000 1010 1020 1030 1040 1050

800 810 820 830 840 850
TACCTCATCGGTCTAGACCACATCCCGCACAGTGCGATTGGTCT--AGAAGATAAT-GCGAGTGCATCACC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
T-GCTCA--GAAGCAGATTCTTCC--CTGAGACCTTCTGACCTCCAAGCCGCTTATAACCTGAGCATCGCC
1060 1070 1080 1090 1100 1110 1120

860 870 880 890 900 910 920
GCTCCTGGGCATCCGTTCTGTGGCAAC---ACGGGCTCTCTTGATCAA-GGCGCTCATCGACCAAGGCTAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
CCTCCT-CAAAT-CTCTCTTGTGACCGTCTTTCGCAATTCCCCCTTCGATGTGCTCTTTGCGCCGAGAT-C
1130 1140 1150 1160 1170 1180

930 940 950 960 970 980 990
ATGA-AACAAATCCTCGTTTTGGAATGACTGGCTGTTTCG--GGTTTTCGA-GCTATGTCACCAACATCATGGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AAGAGGGCATCTCC----CTCCAA-GGGTGTCTTTGCGCTTGATATTGACGCTCCGTG-GCAAGCTC--GCA
1190 1200 1210 1220 1230 1240 1250

1000 1010 1020 1030 1040 1050 1060
CGTGATGGATCGCGTGAACCCCGACGGCA-TGGCCT-TCATTCC--ACTGAGAGTGATCCCATTTCTACGAGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
AGT-ATG---CGCTTG----CCGGCGGCAGTGTGCATCTCGGTCTGACCGAGCCAGA-GTGGTTCAAGGGCA
1260 1270 1280 1290 1300 1310

1070 1080 1090 1100 1110 1120
GAAGGGCGTCCCACAGGAACGCTGCCAGG--CATCACTG--TGACTAACCCG-GCGCGGTTCTGTGTCAACC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCATCGATGACCTCCGTGCTGTCCGTGAGGTCCTTAACGGCATGCCCAACCGGCCCGCGTCTGCG-CAAG
1320 1330 1340 1350 1360 1370 1380

1130 1140 1150 1160 1170 1180
GACTTGCCGTGCATGACGCCATCTGGATCCT----TCCA-----CGCAGCGGCCACTATTCCCC-----G
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
GAGTT--CATCTTTGACGAGTACCAGATCCTTCGAAGCCAGACTTGCCGGTGTGACTGTTCTCCTCATTG
1390 1400 1410 1420 1430 1440 1450

1190 1200 1210 1220 1230 1240 1250
TCAAGATACCGAACGATGAAGTCGCGCATCGATCG-ATAGGCA--TCTTCAATGTGATCAGGGCTGCCACCT
||||||| | | |||| | | | |||| | | | ||| | | |||
TCAAGATGCTCGAGTATGA----GCTCCGCGAGCGCCTATACAAGTACTCCTTGT-CTCTCGGCATGGAGC-
1460 1470 1480 1490 1500 1510 1520

1260 1270 1280 1290 1300 1310
CCAAAGCCGGTGGCC--ACCCCTGTCGATA-GTCTTGAGGGA-CGGTAGCGACGACCGTGCTTTT---CGTG
|| ||||| | | | | | | | | | | | | | | | | |
CCCTAGTCGAGGTCCAGAACACGAGGAGATGGCCACAGCCATCAAGCTCGGCG-CCAAGGTTATCGGCGTC
1530 1540 1550 1560 1570 1580 1590

X 10 20 30 40 50
CTGCAGCCTGACTCGGC-ACCAETCGCTGCAAGCAGAGTCGT--AAGCA-ATCGCAAGGGGG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
GGTTGGCGTCCGCCACC-AAGTCCTCTTTGAGGGGTTCAACTGGGAGTCGTGGAAGCAGAGCGGCGGGTG
160 X 170 180 190 200 210 220

60 70 80 90 100 110 120
--CAGCATGCAAACGAGAAGGGTTGTGCTCAAGTCTGCGGCCGCAGGAAGTC-TGCTGGGCGGCCTGGCTGG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 84, 3112-3116 (1987)
 STANDARD full staff_review
 COMMENT Computer-readable sequence for [1] kindly provided by S. Silver,
 25-MAR-1987.

FEATURES Location/Qualifiers
 CDS <1..359
 /note="mercuric reductase (merA; AA at 3)"
 CDS 374..1012
 /note="organomercurial lyase (merB)"
 CDS 1124..1489
 /note="mercury resistance protein (merD)"

BASE COUNT 391 a 641 c 680 g 441 t
 ORIGIN Unreported.

Initial Score = 95 Optimized Score = 590 Significance = 10.17
 Residue Identity = 50% Matches = 727 Mismatches = 511
 Gaps = 194 Conservative Substitutions = 0

```

      X      10      20      30      40      50
      CTGCAGCCTGACTC-GGCACCAGTCGCTGCAAG-CAG--AGTC-GTAAGCAATCGCA--AGG
      ||      || || || || || || || || || || || || || || || || || || ||
TGGCGGTGTCTTGGTATTGCCGCAGGAAGCAGCCGACGTTCTGTCAGTCCTTCTGTTGCCATGTACATTTCT
      800      X 810      820      830      840      850      860

      60      70      80      90      100      110      120
      GGGCAGC-ATGCAAACG--AGAAGGGTTGTGCTCAAGTCTGCGGCCGCAGGAACTCTGCTGGGCGGCCCTGGC
      ||| || || || || || || || || || || || || || || || || || || || || || ||
      TTGCATCTGTCCCGACCGCGGAAGACTGGGCCTCCAAGCATCAAGGATTGGAA--GGATTGGCGATC-GTC
      870      880      890      900      910      920      930

      130      140      150      160      170      180      190
      TGGGTGCGCGACGTGGCTGGATC-GATCGGCACAGGCGATCGGATCAATACGTG-CGCGTCCTATCACAATC
      || || || || || || || || || || || || || || || || || || || || || || ||
      AGTGT---CCACGAGGCT--TTCGGCTTGGGCCAGGAG-TTTAATCGACATCTGTTGCAGACCATGTC-ATC
      940      950      960      970      980      990

      200      210      220      230      240      250      260
      TCTGA-AGCG-GGTTTTCACACTGACTCAGGAGACATCTGCGGCAGCTCGGCAGGATTC-TTGCCTGC---T
      || || || || || || || || || || || || || || || || || || || || || || ||
      TAGGACACCGTGATCGGATATCGACCCA--ATG--TTCTACGGCACCGGCATCGGATTGCGAGCGCGCGGAT
1000      1010      1020      1030      1040      1050      1060

      270      280      290      300      310      320
      TGGCCAGAGTTCTTCGGTAGCCGCAAAGC--TCTAGCG--GAAAAG--GCTGTGAGAGGATTGCGCGCCAGA
      || || || || || || || || || || || || || || || || || || || || || || ||
      TGAACTCGGCGAAACGGTATATGCATTGCCGTGAACCGACCAAAGGAGGTGTTGATGAACGC-CTACACG
1070      1080      1090      1100      1110      1120      1130

      330      340      350      360      370      380      390
      GCG-GCTGGCGTGCGAACGATTGTGATGTGTGACTTTTGATATCGGTGCGACGTGAGTTTATTGGCCGA
      || || || || || || || || || || || || || || || || || || || || || || ||
      GTGTCCCGGCTGGCCCTTGA-TGCCG--GGGTGAGCGTGC-ATATCGTGCAGGACTAC---CTGCTGCGCG-
1140      1150      1160      1170      1180      1190      1200

      400      410      420      430      440      450      460
      GGTTC-GCGGGCTGCCGACGTTTCATATCGTGGCGGCGACCGGCTTGTGGTTCGAC-CCGCCACTT---TCG
      || || || || || || || || || || || || || || || || || || || || || || ||
      GATTGCTGCGGCCAGTCG-CCTGCACCACG-GGTGGCTA-CGGCCTGTTGATGACGCCGC--CTTGCAGCG
      1210      1220      1230      1240      1250      1260

      470      480      490      500      510      520
      A-TGCGAT---TGAGGTATG---TAGAGGAACCTCACACAGTTCTTCCTGCGTGAGATTCAATATGSCATCGA
      || || || || || || || || || || || || || || || || || || || || || || ||
      ACTGTGCTTCGTGCGGGCCGCCCTTCGAGG--CGGGCATCGGCCT--CGGCG---CATTGGCGCGGCTGTGCC
1270      1280      1290      1300      1310      1320      1330

```

530 540 550 560 570 580 590
AGACACCGGAATTAGGGCGGGC-ATTATCAAGGTCCGACCA-CAGGCAAGGCGACCCCCTTTC-AGGA--G
| | | | | | | | | | | | | | | | | | | | | |
GGGCGCTGGA--TGCGGCGAACTGCGATGAAACTGCCGCGCAGCTTGCTGTGCTGCGTCAGTTTCGTGCAACG
1340 1350 1360 1370 1380 1390 1400

600 610 620 630 640 650 660
TTAGTGTTAA-AGGCEGC GCCCGGG-----CCAGCT TGGCCACC GG GTTCCGGTAACCACTC--ACACGGC
| | | | | | | | | | | | | | | | | | | | | |
CCGGCGCGAAGCGTTGGCCAATCTGGAAGTG CAG-TTGCCCGCC-ATG---CCG---ACCGCGCCGGCACAGC
1410 1420 1430 1440 1450 1460

670 680 690 700 710 720
A-GC-AAGTCAGCGCGATGGTGAG---CGAG-GCAGGCCCGCCATTTTTGAGTCCGAAGC--TTGAGC-----
| | | | | | | | | | | | | | | | | | | | | |
ATGCGGAGAGTTTGCCATGAACAGCCCCGAGCGCATGCCCGC---CGAGACACACAAGCCGTTTACCGGCTA
1470 1480 1490 1500 1510 1520 1530

730 740 750 760 770 780 790
CCTCACGGGTTTG-TATTGGT-CACAGCGATGATACTGACGATTTGAGCTATCTCACCGCCCTGCTGCGCGG
| | | | | | | | | | | | | | | | | | | | | |
CCTGTGGGGTGCGCTGGCGGTGCTCACC--TGTCCTGTC-ATTGCGGATTCTCGCCATTGTGCTGGCCGG
1540 1550 1560 1570 1580 1590 1600

800 810 820 830 840 850 860
ATACCTCATCGGTCTAGACCACATCCCGCACAGTGC GATTG--GTCTAGAAGATAATGCGAGTGCATCACCG
| | | | | | | | | | | | | | | | | | | | | |
CACGAAGGCCGGTGCGTTTCATCGGACAGCACTGGGGTATTGCAGCCCTCACGCTGA-CCG---GCTTGTTTG
1610 1620 1630 1640 1650 1660 1670

870 880 890 900 910 920
CTCCTGGGCATCCGT-TCGTGGCAAACAGGGCTCTC-----TTGATCAAG-GCGCTCATCGACCAAGGCTA
| | | | | | | | | | | | | | | | | | | | | |
-TCCTG----TCTGTGACGCGGCTGCTGCGGGCCTTCAGAGGTGATCATGAGCGCTTCCCAGCCAA--TTG
1680 1690 1700 1710 1720 1730 1740

930 940 950 960 970 980 990
CATGAAACAAATCCTC--GTTTCGAATEACTGGC-TGTTCCGG-TTTTCGAGCTATGTCACCAAC--ATCAT
| | | | | | | | | | | | | | | | | | | | | |
AATG-GACAGTGGCGCAACTGGCGCA-GGCGGGCCGAGCGCGGGCAGCTTGAGCTGCACTACCAGCCGATTGT
1750 1760 1770 1780 1790 1800 1810

1000 1010 1020 1030 1040 1050 1060
GGACGTGATGGATCGCG-TGAACCCCGACG-GGATGGCCTTCATTCCACTGAGAGTGATCCCATTTCTACGAG
| | | | | | | | | | | | | | | | | | | | | |
CGATTTG-CGCAGTGAGCAGATTGTGCGGCGCGAAGCCCT--GTTGCGCTG-GCGTCATCCGACGCTCGGAC
1820 1830 1840 1850 1860 1870

1070 1080 1090 1100 1110 1120
AGAAG--GGCGTCCACAGCAACGCG-TG--CCAGGCATC---ACTG-TGAC-TAACCCGGCGC---GGTTC
| | | | | | | | | | | | | | | | | | | | | |
TGTTGCCGCGGGGCCAGTTCTGCCCCTGATCGAATCGTCCGGCCTGATGCCGGAAATCGGCGCATGGGTGC
1880 1890 1900 1910 1920 1930 1940 1950

1130 1140 1150 1160 1170 1180
TGTGTCACCGACTTGCGGTGCATGACGCCATCTGGATCCTTCCACGCAGCGGCCACTATTCC-----CCGT
| | | | | | | | | | | | | | | | | | | | | |
TG-GGCGCAG-CCTGCCGTCAAATGCGCGA-CTGG-CGGGTGCTGGCA-TGGCAACCGTTCCGGCTGGCCGT
1960 1970 1980 1990 2000 2010

1190 1200 1210 1220 1230 1240 1250
CAAGATACCGAACGA-TGAAGTCGCGCATCGATCGATAGGCATCTTCAATGTGATCAGGGC-TGCCACCTCC
| | | | | | | | | | | | | | | | | | | | | |
CAATGTTTCG-GCGAGCCAAGTGGGEC-CAGAT--TTCGACA-----AGTGGGTAAAGGGCGTGC--TGCC
2020 2030 2040 2050 2060 2070

TGCACTCCGT--CTCCAACGAAGTTAATGCCTGTCGG--CTGTACGACATCGACGCGTC-CAAGTACG
300 310 320 330 340 350 360

200 210 220 230 240 250 260
GAAGCGGGTTTTCACACTGA-CTCAGG-AGGACATCTGCG-GCAGCTCGGCAGGATTCTTG-C-GTGCTTGGCC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCAACGCG-----GCCGAGCTCAAAGTCGCTCATCGCGCGCTCCACGGCAAG-GGCGTGCAGGCCATCGCC
370 380 390 400 410 420

270 280 290 300 310 320
AGAGTTCTTCGGTAGCCGCAAAGCTCTAGCG--GAAAAGGCTGTGAGAGG-AT-TGC-GC----GCCAGAGC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
-GACATCGTCATCAAC--CACCGCTGCGCCGACTACAAGGATAGCCGCGGCATCTACTGCATCTTCGAGGGC
430 440 450 460 470 480 490

330 340 350 360 370 380 390
GGCTGGCGTGCBAACGATTG--TCGA-TGTGTGACTTTTCGATATCGGTGCGGACGTCAGTTTATTGGCCGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
GGC--ACCTCCG-ACGGCCGCCTCGACTGGGGC-CCCCACATGATCTGTGCGGACGACACCAATACTCCGA
500 510 520 530 540 550 560

400 410 420 430 440 450 460
GGTTTCGCGGGCTGCCGACGTTTCATATCGTGGCGGCGAC--CGGCTTGTGGTTCGAC-CCG-CCACTTTCGA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
---TGGC--ACCGCAACCTCGACACC--GGAGCCGACTTCGCCGCGCGGCCCGACATCGACCAC-CTCAA
570 580 590 600 610 620

470 480 490 500 510 520
TGCGATTGAGGTATGTAGAG-GAACTCACACAGT---TCTTCCTGCGTGAGATTCAATATGGCATCGA----
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
--CGACCG-GGT--CCAGCGCGAGCTCAAGGAGTGGCTCCTCTGGCTCAAGAGCGACCTCGGCTTCGACGCG
630 640 650 660 670 680 690

530 540 550 560 570 580 590
AGACACC-GGAATTAGGGCGGGCATTATC-AAGGTGCGGACCACAG-GCAAGGCGAC-CCCCTTTCAGGAG-
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
TGGCGCCTTGACTTCGCTAGGGGCTACTCGCCGAGATGGCCAAGGTGTACATCGACGGCACATCCCCGAGC
700 710 720 730 740 750 760

600 610 620 630 640 650 660
TTAGTGTTAAAGGCGGCCCGCCCGGSCCAGCTTGGCCACCGG-TGTTCCGGTAACCACTCACACG-GCAGCAA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
CTCGCCGT--GGCCGAGGTGTGGGACAATATGGCCACCGGCGGCGACGGCAAGCCCACTACGACCAGGAC
770 780 790 800 810 820 830

670 680 690 700 710 720 730
G-TCAGCG--CGATGGTGAGCGAGGCGAGGCGCCATTTTGGSTCOGAAGCTTGAGCCCTCACGGGTTTGTA
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCGCACCGGCAGAATCTG-GTGAAGTGGGTGGACAAGGTGGGCGGCGCGGCCCTCGGCAGGCATGGTGTTTGA
840 850 860 870 880 890 900

740 750 760 770 780 790 800
TTGGTCACAGCGATGATACTGACGATTTTGA--GCTATC-TCACCGCCCTGCTGCGCGGATACCTCATCGGTC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
CT--TCA---CGACCAAAGGGA--TTCTGAAGCTGCCGTGGAGGGCGAGCT--GTGGA-GGCTGATCGACC
910 920 930 940 950 960 970

810 820 830 840 850 860 870
TAGACCACATCCCGCACAGTGCATTTGGTCTAGAAGATAATGCGAGTGCATCACCGCTCCTGGGCATC--CG
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
CGCAGGGGAAGGCCCGCGGCTGA--TGAGATCG-TGSCCGGCCAAGGCCCGCCACC-TTCGTCGACAACCACG
980 990 1000 1010 1020 1030

880 890 900 910 920 930
TTCGTGGC-AAACACGGGCTCTCTTGATCAAGGCGCTCATCGACCAAGGC-TACA-TGAAAC----AAATCC
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

ATACAGGCTCCAGCGAGGC-ATTGTGGATTTCCCCTC--CGACAAGTTCATGCGAGGCTACGCGTACATCC
1040 1050 1060 1070 1080 1090 1100

940 950 960 970 980 990 1000
TCGTTTTCGAATGACTGAGCTGTTCCGGTT-TTCGAGCTATGTCACCAACATCATGGACG-TGATGGATCGCGT
|| || | || || || || || ||
TCACCCACCCCGGCATCCCATGCATCTTCTACGA-CCATTTCCTCAAC----TGCGGGGTTTTAAGGA---CCA
1110 1120 1130 1140 1150 1160 1170

1010 1020 1030 1040 1050 1060 1070
GAACCCCGACGGGATGACCCTTCATTCCACTGAGAGTGATCCCATTTCTACGAG-----AGAAGGGCGTCCAC
|| || || || || || || || ||
GATCGCGGCGCTGGTGGCGATCAGGAAGCGCAACGGCATCAC-GGCGACGAGCGCTCTGAAGATCCTCATGC
1180 1190 1200 1210 1220 1230 1240

1080 1090 1100 1110 1120 1130 1140
AGGAA-ACGCTGCCAGGCATCACTGTGACTAACCCGGCGCGGTTCTGTGTACCGACTTGCCGTGCATG-AC
| || || || || || || || ||
ACGAAGGAGATGCC-TACGTGCGCCGAGA-TA-GACGGCAAGGTGGTG-GTGA-AGA-TCG-GGTCCAGGTAC
1250 1260 1270 1280 1290 1300

1150 1160 1170 1180 1190 1200 1210
GCCATCTGGATCCT--TCCACGCAAGCGGCCACTATTCCCCG--TCAAGATACCGAACGATGAAGTC-GCGCA
| || || || || || || || ||
GACGTGCGGGGCGGTGATCCCGGC--CGGGTTCTGTGACCTCGGCACACGGCAACG-ACTACGCCGTCTGGGAG
1310 1320 1330 1340 1350 1360 1370

1220 1230 1240 1250 1260 1270
TCGATCGAT----AGGCA--TCTTCAA-TGTGATCAGGGCTGC-CACCTCC-AAAGCCGGT-GGCCACCCCT
|| || || || || || || ||
AAGAACGGTGCCGCGGCAACACTACAACGGAGCTGAAGTCTGCACTGATCCGTCAATTCGATCGAGCATGAAT
1380 1390 1400 1410 1420 1430 1440

1280 1290 1300 1310 1320 X
GTC--GATAGT-CTTGAGGGAC----GGTAGCGACGACCG--TGCTTTTC-GTGAAGTGCAG
|| || || || || || || ||
TTCCTGA-AGTACATGATTCACTTCTGGTATTACG-CGGATATGATTAAGTATGTATACCTGTACCCAAAA
1450 1460 1470 1480 1490 1500 1510

T
1520

(440 aa).

EMBL features not translated to GenBank features:

key	from	to	description
RBS	344	347	put. rRNA-binding site

FEATURES

	Location/Qualifiers
CDS	<1..353
	/note="cyaD polypeptide (AA at 3)"
CDS	355..1779
	/note="cyaE polypeptide (AA 1-474)"

BASE COUNT 282 a 679 c 770 g 309 t

ORIGIN

Initial Score	=	82	Optimized Score	=	590	Significance	=	10.17
Residue Identity	=	51%	Matches	=	732	Mismatches	=	495
Gaps	=	197	Conservative Substitutions	=			=	0

```
X          10          20          30          40          50          60
CTGCAGCCTG-ACTCGGCACCAAGTC---GCT-GCAAGCAGAGTCGTAAGCAATCGCAAGG-GGGC-AGCAT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCGGC-GCCGGCATCCAGGTCCAGGCTCAGCTCGACAGCAAGGACATCGGCTTTGTTCAGGGCGGGCGCGCCA
X          10          20          30          40          50          60          70
      70          80          90          100          110          120          130
GCAAACGAGAAGGGTTGTGCTCAAGTCTGCGGCCGCAGGAAGTCTGCTGGGCGGCCTGGCTGGGTGCGCGAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCTACCGTCAA-GGTGG-GC----GCCTACGACTATACGAAGTACGGAACGCTCGAAGGCAAGGTGTTGTAT
      80          90          100          110          120          130
      140          150          160          170          180          190          200
GTGGCT--GGAT-CGATCGGCACAGGCGATCGGATCAATACGTGCGCGT--CCTATCACAACTCTCTGAAGCG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTGTCTCCGGATACGGTGGTC---GACGACCG--CCAACA-GCACTCGTACCGCGTGACGATCGC----GCT
      140          150          160          170          180          190
      210          220          230          240          250          260
GGTTTCACACTGAC--TCACGAGGAC----ATCTGCGGCAGCTC--GGCAGGATTCTTGCGTGCTTGGCCAG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GG-CGCACCCTGCCCTGGAGGTGGACGGCAAGCCGCGGCTGCTCAAGGAAGGCATGGCG-GTGC-AGGCC-G
200          210          220          230          240          250          260
      270          280          290          300          310          320          330
AGTTCTTCGG--TAGC-CGCAAAGCTCTAGCG-GAAAAGGCTGTGAGAGGATTGCGCGCCAGAGCGGC--TG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
A--TATCCGGACCGGCTCGCGGCGCCTCATCGAGTATCTGCTCAGCCCGG-TGGCGCGGCATGCCGGCGAAA
      270          280          290          300          310          320          330
      340          350          360          370          380          390          400
GC--GTGCGAACGATTGTGATGTGTGACTTTGATATCGGTGCGGACGTCAGTTTATTGGCCGAGGTTTC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GCCTGGGGGAGCGCTAG--CATG-GCCG-CGGTCAAGTCAGGCGACGCGGCCGGGCCCTGGCGTTGGCGCT
      340          350          360          370          380          390          400
      410          420          430          440          450          460
GCGGGCTGCCGACGTTTCATATCGT-GGCGGCG----ACCGGCTTGTGGTTTCGACCCGCCACTTTTCG-ATGCG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GTGGGCCGGGTTTCGCGCTGAGCGTGGGAGGCGGGGTGCGGCGCGCGAT--GGCCTGGCAACGCCGCCCGCG
      410          420          430          440          450          460          470
470          480          490          500          510          520          530
ATTGAGG--TATGTAGAG---GAACTCACACAG--TTCTTC-CTGCGTGAGATTCAATATGGCATCGAAGAC
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TTCGAGGGCCAGGCGGCGCCTGCCGTCTCGTGGCCTTGTCCGCCGCCGGCGGATC-----GGC-TCGACGAC
      480          490          500          510          520          530          540
      540          550          560          570          580          590          600
ACCGGAATTAGGGCG-GGCATTATCAAGGTCCG-----ACCACAGGCAAGGCGACCCCCCTTTCAGGAGTTAG
```